IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 December 2023, accepted 11 December 2023, date of publication 19 December 2023,
date of current version 27 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3344161

INVITED PAPER

Noise Driven Coupled Nonlinear Systems: A "Dynamical Multilayer Perceptron" Demonstrating XOR Functionality

PAUL R. DE LA HOUSSAYE[1], (Senior Member, IEEE), JUSTIN MAUGER[1], ADI R. BULSARA[1], CHRISWELL HUTCHENS[2], (Life Member, IEEE), AND BENJAMIN MIGLIORI[3], (Member, IEEE)

[1]Naval Information Warfare Center Pacific, San Diego, CA 92152, USA
[2]Department of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA
[3]Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Corresponding author: Paul de la Houssaye (paul.r.delahoussaye.civ@us.navy.mil)

This work was supported in part by the Naval Innovative Science and Engineering (NISE) Program, Naval Information Warfare Center (NIWC) Pacific; and in part by the Internal Seedling Grant, NIWC.

ABSTRACT Coupled nonlinear dynamical systems are known to display remarkable behavioral similarities to some aspects of neural activity in biology. When considering the computational properties of such a neuromorphic network (or even an electronic biomimetic computation device underpinned by such a network), the question of whether (and how) such a network realizes fundamental logic operations must be examined. In this work, a dynamical system of coupled noisy overdamped nonlinear bistable elements is considered. The background noise floor is used to drive information flow by helping each element switch randomly between its stable states, with the system response quantified via a long-time probability density function. A theoretical representation of such a system is developed and a simple five-element realization is used to demonstrate a continuous version of an XOR gate, through the proper choice of coupling coefficients and controllable external biases. This is a first step towards synthesizing more general functions, a prerequisite for advanced computing and learning applications. Finally, a silicon implementation is proposed and simulated using a verified process model that could be fabricated to generate a working analog computing system.

INDEX TERMS Dynamical nonlinear oscillator, network, noise-floor, stochastic differential equation, XOR gate.

I. INTRODUCTION

Artificial Neural Networks (ANNs) [9] are ubiquitous in increasingly sophisticated scenarios in the general areas of computer vision, natural language processing, and neuromorphic computing. Such systems afford capabilities beyond traditional methods, and they hold out the tantalizing promise of more elegant and efficient applications in areas such as robotics and artificial intelligence. There also exist a number of dynamical mathematical models that mimic certain aspects (e.g. spiking, long term potentiation) of biological neural networks (see e.g. [4]). Such models can facilitate the modeling of the interaction between the cell body and synapses, allow one to explore spatiotemporal complexity, and provide a framework to examine the effects of background noise. This dovetails with recent interest in collective behavior in large scale brain models, as a possible driver of brain functions, e.g., cognition and perception [1], [2], [3]. However, these models can be limited in terms of algorithmic implementation, as they evolve towards larger and more complex networks. In recent years, many-core neuromorphic processors [5] as well as Very Large-Scale Integration (VLSI) emulations of large neural networks with as many as 10^6 silicon neurons [6] have been developed. In this work, we present a potential method for driving neural-like dynamics by taking advantage of the inherent

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

(thermal or other) noise in networks of coupled nonlinear dynamic elements, each with a noise floor; this results in sufficiently complex behavior to perform more advanced tasks such as function approximation.

As a necessary prerequisite for universal function approximation, artificial neural networks should be able to reproduce the basic logic gate functions. The AND and OR gates can each be implemented by a single layer network consisting of two input elements and one output element. In contrast, implementing the XOR gate requires a two layer network, with an additional two elements in a "hidden" layer. We focus on implementing an XOR function, operating on analog variables, due to its ubiquity in the machine learning literature. This can be viewed as a first step towards synthesizing more complicated functions.

Here, we examine how noise-driven coupled nonlinear dynamical systems can yield XOR functionality. We consider a network of bistable nonlinear dynamic elements with nonlinear coupling and a noise-floor attributed to each element. These elements qualitatively mimic the dynamic behavior of neurons, with the important caveat that they do not produce spiking; the bistability (underpinned by a nonlinear potential energy function) occurs between two fixed points, rather than between a fixed point and a limit cycle. This system provides a reasonable model for describing, in the mean, the interaction between the noise floor and the stimulus in the network; the deterministic solution is supplanted by a statistical solution wherein the *averaged* response can be calculated, and connected to observables in the real system. Our proposed system is inspired by conventional ANNs, which may be viewed as consisting of *static* coupled nonlinear elements. Our choice of nonlinearity closely mimics the usual sigmoidal transfer characteristic commonly used in ANNs, with the coupling also being a sigmoidal (hence bounded) function with suitably chosen coefficients or "strengths".

Our dynamical network also resembles a "liquid state machine" (LSM), which has been proposed for computing with time series [10]. Like LSMs, our network can handle time-dependent inputs, and offers the possibility of quadrature via techniques involving time scale separation. Furthermore, although we take the coupling coefficients to be constants in this study, they can readily be described statistically for more complex scenarios. Nonlinear dynamic elements comprising LSMs can also exhibit cooperative dynamical behavior arising from the interplay between nonlinearity, coupling, and noise, most notably when cast in the framework of driven dynamical systems. One example of such behavior is the oft-studied Stochastic Resonance (SR) effect [11], which is known to be amplified in a nonlinear sensor network via the interplay of coupling, applied signal (often taken to be time-periodic), and noise floor [12], [13]. Of particular interest in the context of this paper, we note the "Logical Stochastic Resonance" (LSR) [14], [15], [16] scenario, wherein a noisy bistable element accepting two random data streams can reproduce the fundamental logic operations (AND/OR, NAND/NOR) and morph from one to the other using a controllable dc bias to perturb the underlying potential energy function; the gate probability is unity over a definite regime of noise intensity. Notably, a possible biological realization of LSR was proposed by Dari et al. [17], [18] in a theoretical description of an engineered gene network.

In this work, all the computational elements are dynamic in nature, including the output node; each nonlinear element is characterized via a stochastic differential equation with nonlinear inter-element coupling terms, and subject to a deterministic (taken to be constant in this work) applied bias signal. The network elements switch between their stable steady states in the presence of noise with an appropriate selection of coupling strengths. The output of the system will thus not be deterministic, and any result stemming from the network has a probabilistic structure from which statistical measures, such as mean values can be computed, via a long-time probability density function. In Section II, we construct an array consisting of a single overdamped bistable "reference" element coupled to an ensemble of "bath" elements which evolve on significantly faster time scales than the reference element. Hence, the long-time system behavior is well approximated by the dynamics of the reference (or "slow") element. It is assumed that the bath is never far from its long-term steady state. Under these assumptions, the dynamics of the reference element can be obtained in closed form using an adiabatic elimination technique or "slaving principle" first propounded by Haken [19]. The response (for assorted constant input signals and noise intensities) of the network is quantified by the steady-state (i.e. at long time) *probability density function* of the reference element. As an aside, we note the similarity of our system to "liquid networks" [20]. Analogous to liquid networks, our noisy system yields a random 2-state time series solution. However, the relevant information for realizing the XOR is contained in the long-time probability density function $P(y_1)$ of the readout element, rather than the actual time-series solution.

We consider, in this work, the particular case of a small network of five elements, based on the architecture of a particular ANN which implements the XOR. The network consists of two input elements (each of which accepts an input dc signal) coupled to a single output or readout element while passing through an intermediate or "hidden" layer comprising two elements. Each element has its own noise floor (assumed to be uncorrelated from element to element). Hence, our system implements an analog transform that approximates the XOR function *in probability space*. The transform has analog inputs, whereas the output consists of probability distributions that can be partitioned into two equivalence classes representing a "TRUE" or "FALSE" result based upon whether the dc inputs have opposite signs, or the same signs, respectively. We note that the theory derived in Section II is valid for networks of arbitrarily sized inputs and hidden layers and involves forward and backward coupling, subject to certain constraints on the parameters.

In Section II, we introduce the network via a system of coupled stochastic differential equations (SDEs). A Fokker Planck Equation (FPE) for the network is set up, then decoupled using the above-mentioned adiabatic approximation. The FPE is a diffusion type equation that describes the evolution of the ($N-$ body) probability density function; it involves averaged quantities such as the mean and variance of the noise in each element, rather than the noise itself. The FPE-based approach is particularly advantageous for the network under consideration; we will touch on this point later in greater detail. Eventually, we are led to a single effective SDE for the readout element (this is the "slow" element in the framework of the adiabatic approximation), which is the goal of the theory. From this SDE, we readily derive a closed form probability density function for the readout element in the long time limit; we will see that the FPE description coupled with the "slaving" argument accomplishes this goal quite neatly. The effects of the remaining "fast" elements in the network are manifested in a skewing of the slow element probability density function; effectively the fast elements constitute a "bath" to which the slow (i.e. readout) element is coupled. It will, additionally, become clear that our general procedure based on a decoupling of the many-body FPE can be readily applied to larger networks and even more complex coupling schemes.

In Section III-A, the results obtained via the analytic model of our network are compared with numerical simulations and shown to match within the constraints of our approximations; these constraints will be defined as we proceed with the calculations. Finally, in Section IV, we consider the implementation of the coupled dynamical system in hardware by carrying out circuit simulations using a validated Process Design Kit (PDK). We demonstrate that the circuit implementation yields nonlinear dynamic behavior (in this case hysteresis) in the output (or readout) element, which matches the predictions of the theoretical model remarkably well. While we do not explicitly use the nonlinear circuit to realize the XOR (this is left to an upcoming paper), the agreement of the theory with the circuit realization gives us the confidence to use it as a roadmap to adjust the circuit parameters to obtain the desired dynamical behavior. This is especially important as the number of elements and/or complexity of the system grows, and direct simulation becomes time prohibitive.

II. THE COUPLED DYNAMICAL NETWORK

We consider a network of coupled overdamped bistable elements, each subject to noise and a constant deterministic signal. The choice of nonlinearity in our dynamics is important. We use a hyperbolic tangent function that is bounded above and below and, hence, reproduces the sigmoidal transfer function that historically has been used to mimic the on-off dynamics of biological neurons.

In its most general form, the network can be described by the coupled system of stochastic differential equations for a generic state variable $y(t)$ (e.g. in a nonlinear circuit, as considered later in this work, $y_i(t)$ represents the voltage in the i^{th} element)

$$\dot{y}_i = q_i(\{y_j\}) + N_i(t) \tag{1}$$

where the overdot denotes the time-derivative. The noise $N_i(t)$, in the i^{th} element, is assumed to be Gaussian, and delta-correlated with zero mean and variance σ_i^2; further the noise sources in different elements are assumed uncorrelated. We note that this so-called "white" noise is an idealization. Rigorously, $N_i(t)$ should be correlated (or "colored") noise obtained via an Ornstein-Uhlenbeck process [23], [24] that leads to exponentially correlated noise with correlation time τ_c which is, roughly, the inverse of the noise bandwidth:

$$\langle N_i(t)N_i(s)\rangle = \frac{\sigma_i^2}{2\tau_c}\exp\Big[-\frac{\mid t-s\mid}{\tau_c}\Big] \tag{2}$$

which corresponds to a power spectral density

$$S_i(\omega) = \frac{\sigma_i^2}{1+\omega^2\tau_c^2}. \tag{3}$$

In the limit $\tau_c \to 0$, the rhs of (2) becomes a δ function i.e. we obtain Gaussian $\delta-$ correlated noise, having (by definition) zero mean:

$$\langle N_i(t)\rangle = 0, \quad \langle N_i(t)N_j(t')\rangle = \sigma_i\delta_{ij}\delta(t-t'). \tag{4}$$

The power spectral density, in this "white" limit, becomes a constant equal to the noise variance σ_i^2. We point out that having the noise autocorrelation function (in the colored case and the corresponding white limit) that is a function only of the time difference, is a hallmark of a Markov process [23].

As noted above, the (theoretical) concept of infinite bandwidth or "white" noise is unrealistic and unrealizable in experiments; the noise can usually be adjusted to have very large (however, not infinite) bandwidth. In practice, it is sufficient that the noise have a bandwidth much greater than other system frequencies of interest (e.g. the inverse of the time constants a_i and any externally applied frequencies), in order to use the white noise representation (4). We note that the noise arising in real systems can have a wide range of bandwidth, arising from the underlying physics. As an example, Superconducting Quantum Interference Devices (SQUIDS) [28] typically have a noise floor with gigahertz bandwidth, whereas noise in some biological systems might have a bandwidth of only a few hertz. In both these cases the noise may be taken to be broadband; what matters is for the bandwidth to be larger than any other frequency (or inverse time) in the system. In Section III we will detail how the noise bandwidth and the simulation time-step are intertwined.

We observe that, in our notation employed in (1), the noise $N_i(t)$ has unit variance and is multiplied by the standard deviation σ_i. The rhs of (1) is

$$q_i(\{y_j\}) = -a_i y_i + \sum_{j=1}^{N} J_{ij} \tanh y_j + \varepsilon_i \tag{5}$$

where each element is subject to a constant bias term ε_i. The nonlinear terms include the "self" coupling with strength J_{ii}, as well as nonlinear coupling (via a coupling coefficient $J_{ij}, j \neq i$) to the other elements. The coefficients a_i have units of inverse time and must be positive to maintain dynamic stability.

A. BACKGROUND: SINGLE ISOLATED ELEMENT

We begin with a concise dynamical description of a single uncoupled (i.e. isolated) element, which is the basic building block of our network. Its dynamics are (we drop the subscripts for convenience)

$$\begin{aligned}\dot{y} &= -ay + J\tanh y + \varepsilon + N_1(t)\\ &\equiv -\frac{\partial}{\partial y}U(y) + N(t),\end{aligned} \tag{6}$$

with the potential energy function given by

$$U(y) = \frac{a}{2}y^2 - J\ln\cosh y - \varepsilon y, \tag{7}$$

and the "switching curve" corresponding to the deterministic portion of the rhs of (6)

$$\frac{\partial}{\partial y}U(y) = 0 = ay - \varepsilon - J\tanh y. \tag{8}$$

Clearly, in the presence of additive (Langevin) noise, the extrema of the potential energy function are obtained using the *deterministic* dynamics. The noise manifests itself when we consider the probability density function of element y. This probability density function can be obtained by setting up the Fokker Planck Equation (FPE) that governs its temporal evolution: [23]:

$$\frac{\partial P(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left[A(y)P(y,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 P(y,t)}{\partial y^2}, \tag{9}$$

with the drift (or streaming) term having the kernel $A(y) = -\partial U(y)/\partial y$. The important thing to note is that the FPE contains the noise dependence in a second order diffusion term involving the variance σ^2. In general, solving the FPE can be quite daunting, except when the kernel $A(y)$ is linear or possessed of some other convenient functional form [23]. However, *in the absence of an explicit time dependence on the rhs of (9)*, one may write down the long time or steady-state solution to (9) by setting its lhs to zero, and integrating the resulting differential equation [23]. The result is:

$$P(y) = \mathcal{N}\exp\left[-\frac{2}{\sigma^2}U(y)\right] \tag{10}$$

where $\mathcal{N}$ is a normalization constant. This indicates that the spread of the probability density function is directly dependent on the noise variance. We note that, for a more complicated case wherein there is an explicit time-dependence on the rhs of (9), one cannot simply write down a solution of the form (10); it would be necessary to obtain the full time dependent solution to (9), and then find its limit as $t \to \infty$. In simulations, one would assemble a time-series of the stochastic variable $y(t)$. Depending on the choice of parameters this series displays a single state or bistable state behavior, corresponding to a potential energy function that is monostable or bistable. All the important system characteristics are manifested in the potential energy function (7) wherein the parameters (a, J) are obtained from the physics of the particular system under consideration. Knowing this potential energy function is the critical stepping stone to understanding the system behavior.

The locations of the extrema of the potential energy function, for the bistable case, can be determined using simple calculus. First, we consider the symmetric case $(\varepsilon = 0)$. The extrema of the potential energy are given by the roots of $y - \eta\tanh y = 0$ (where $\eta \equiv J/a$), the rhs of the (previously defined) switching curve. The trivial solution is $y = 0$, which can be shown to correspond to the unstable fixed point (maximum) of the potential function. We readily observe that $\eta > 1$ is the *condition for bistability*. For $\eta \leq 1$ the potential function becomes monostable, however it becomes perfectly parabolic only in the $J \to 0$ limit. The remaining extrema are found by approximately obtaining the (non-zero) roots of the switching curve. We find for the (stable) minima $y_\pm(\varepsilon = 0)$:

$$y_\pm(\varepsilon = 0) = \pm\eta\left[1 - \frac{1 - \tanh\eta}{1 - \eta\,\mathrm{sech}^2\,\eta}\right] \approx \pm\eta\tanh\eta \tag{11}$$

where the approximation becomes exact for large η, and the unstable fixed point is, as noted above, at $y = 0$. For non-zero ε the minima are shifted and a simple expansion yields

$$y_\pm(\varepsilon) = y_\pm(\varepsilon = 0) + \delta; \quad \delta = \frac{\varepsilon}{1 - \eta\,\mathrm{sech}^2\, y_\pm(\varepsilon = 0)}, \tag{12}$$

which can be simply written as:

$$y_\pm(\varepsilon) = \pm\eta\tanh\eta + \frac{\varepsilon}{1 - \eta\,\mathrm{sech}^2(\eta\tanh\eta)} \tag{13}$$

For this case $(\varepsilon \neq 0)$, the unstable fixed point is, of course, located at $y = \delta$. Implicit in the derivation of (13) is the assumption that $\varepsilon/J \ll 1$. The last statement requires qualification. If ε is too large, it will render the potential monostable. In this case the three roots of the switching curve become a single real root and two complex conjugate roots. Hence, if we require the potential to remain bistable (as throughout this paper), this constrains the dc signal ε.

B. FEED-FORWARD NETWORK

1) STATIC XOR

In this work, we wish to create a network that replicates the behavior of an XOR gate. At its core, the XOR gate is a function of two Boolean variables returning FALSE if the input variables are the same, and TRUE if the input variables are different. If we encode TRUE as $+1$ and FALSE as -1, we can define an analog version of an XOR gate acting on real-valued variables as any continuous function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ satisfying

$$\begin{aligned} f(1,1) &= f(-1,-1) = -1 \\ f(1,-1) &= f(-1,1) = 1. \end{aligned} \tag{14}$$

We also add the requirement $f(x,y) \le 0$ if $xy \le 0$ and $f(x,y) \ge 0$ if $xy \ge 0$. One particular ANN that approximates these requirements is the two layer network with functional form

$$f_\beta(x,y) = \tanh\big(\beta(1+x+y)\big) + \tanh\big(\beta(1-x-y)\big) \tag{15}$$

wherein the error–defined as the discrepancy between f_β and f in (14)–can be made arbitrarily small by increasing β. The decision boundary of this network (the zero-locus of f_β) can be made arbitrarily close to the pair of parallel lines $x+y=\pm 1$. The implementation of the XOR in an ANN is described in [21] and [22].

In the next section, we create a network of coupled dynamical oscillators with the same architecture as this ANN, and with parameters guided by (15). This network will be the underpinning of a "dynamic XOR". We also demonstrate that $N = 5$ elements is the minimum configuration required to implement the XOR, similar to the case with ANNs.

2) DYNAMIC XOR

Based on the discussion above, we set up a coupled dynamical system of five elements in a two-layer configuration (see Figure 1). For simplicity, we do not include backcoupling, making this a *feed-forward* network. We start from our general network equation (1) and write out the q_i:

$$\begin{aligned} q_1 &= -a_1 y_1 + J_{11}\tanh y_1 + J_{12}\tanh y_2 + J_{13}\tanh y_3 + \varepsilon_1 \\ q_2 &= -a_2 y_2 + J_{22}\tanh y_2 + J_{24}\tanh y_4 + J_{25}\tanh y_5 + \varepsilon_2 \\ q_3 &= -a_3 y_3 + J_{33}\tanh y_3 + J_{34}\tanh y_4 + J_{35}\tanh y_5 + \varepsilon_3 \\ q_4 &= -a_4 y_4 + J_{44}\tanh y_4 + \varepsilon_4 \\ q_5 &= -a_5 y_5 + J_{55}\tanh y_5 + \varepsilon_5. \end{aligned} \tag{16}$$

The structure of these terms describes an input layer comprised of 2 elements $y_{4,5}$, with dc inputs $\varepsilon_{4,5}$ and (uncorrelated) noise terms $N_{4,5}(t)$. The intermediate layer, comprising elements $y_{2,3}$ with analogous dc and noise driving terms, was coupled to the input layer. Finally, the (single) readout element y_1 draws its input from the intermediate layer.

System (1) with definitions (16) cannot be solved analytically in the presence of noise sources unless appropriate approximations are made. We follow our previous work [12] and assume that the "readout" element y_1 evolves far slower than the remaining elements, thereby allowing us to look at the problem as a system (the "slow" element y_1) coupled to a "bath" of (in this case) 4 elements, with the latter reaching their steady states far more rapidly than the reference or readout element y_1. The conditions for this to occur are detailed in [12] and [19]. For our purposes the condition can be boiled down to demanding that $a_1 \ll a_i, i = 2\ldots5$. We demand that each individual element is bistable when isolated, meaning that $J_{ii}/a_i > 1\ \forall i$. This sets us up to consider the most interesting case while realizing that the coupling and dc forcing terms can render one or more elements monostable in the (coupled) network. An analytic solution of the coupled stochastic differential equations (SDEs) (1) would be impossible under any other scenario; however, using the above-mentioned adiabatic approximation [19], we aim to obtain the long-time probability density function $P(y_1, t \to \infty)$ which we denote as $P(y_1)$ for convenience. The structure of $P(y_1)$ under assorted coupling, noise and forcing scenarios will inform the realization of the XOR gate.

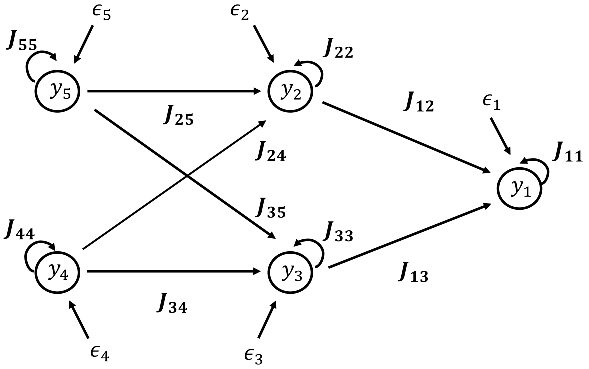


FIGURE 1. A schematic representation of a 5-element, 2-layer system. J_{ik} are the coupling coefficients between the individual elements, whereas $\varepsilon_{i=1-5}$ are the (controllable) constant (i.e. dc) biases applied to each element. $\varepsilon_{4,5}$ are, therefore, dc inputs to the system. y_1 is a "slow" element, whereas $y_{i=2-5}$ are "fast" elements (see text). The steady state probability distribution of the y_1 element is taken as the relevant output of the system. $y_{2,3}$ constitute an "intermediate layer" which is subject to dc bias ($\varepsilon_{2,3}$) and noise ($N_{2,3}$). For simplicity, this paper focuses only on a feed-forward network with all back-couplings ($J_{21}, J_{31}, J_{42}, J_{52}, J_{43}, J_{53}$) set to zero.

The starting point is the N-body Fokker-Planck equation (FPE) for the probability density function $\mathbf{P}(\{y_i\}, t)$ corresponding to the system of variables y_i (corresponding to the dynamically evolving nodes in the network of Figure 1) that satisfy the coupled SDEs (1). This may be written as [23]

$$\frac{\partial \mathbf{P}(\{y_i\},t)}{\partial t} = \sum_{i=1}^{N}\left[-\frac{\partial}{\partial y_i}(q_i\mathbf{P}) + \frac{\sigma_i^2}{2}\sum_j \frac{\partial^2 \mathbf{P}}{\partial y_i \partial y_j}\right] \tag{17}$$

with the drift terms defined in (16) and the noise terms manifesting themselves in the second order diffusion terms. The "slaving" argument [19] then leads us to factor the probability $\mathbf{P}(\{y_i\}, t)$ as follows

$$\mathbf{P}(y_1, y_2, \ldots, y_N) = h(y_2, y_3, \ldots, y_N \mid y_1)P(y_1) \tag{18}$$

(suppressing t to simplify our notation) where h is to be interpreted as a conditional probability density for finding $y_{j>1}$ given y_1 (both h and P are normalized to unity). The slaving constraints (i.e. the separation of time-scales embodied in the relative values of the coefficients a_1 and $a_{i>1}$) then imply that we may substitute the factored probability density function (18) into the original FPE (17) and separate the slow variable y_1 from the fast one(s) [12], [19]. After some algebra, we obtain

$$\frac{\partial h(y_{i>1},t)}{\partial t} = \sum_{i>1}\left[-\frac{\partial}{\partial y_i}(q_i h) + \sum_{j>1}\frac{\sigma_i^2}{2}\frac{\partial^2 h}{\partial y_i \partial y_j}\right] \tag{19}$$

$$\frac{\partial P(y_1,t)}{\partial t} = -\frac{\partial}{\partial y_1}\left[A(y_1)P(y_1)\right] + \frac{\sigma_1^2}{2}\frac{\partial^2 P}{\partial y_1^2} \tag{20}$$

with the kernel $A(y_1)$ defined by

$$A(y_1) \equiv \int \ldots \int h(y_2, y_3, \ldots, y_N \mid y_1) \times q_1(y_1, y_2, \ldots, y_N)dy_2dy_3\ldots dy_N. \tag{21}$$

The above equation represents the formal (N-element) decoupling of the Fokker-Plank equation with the probability density function $P(y_1,t)$ representing the slow element dynamics. We have assumed, in deriving the above (decoupled) FPEs that the derivative of h with respect to y_1 may be neglected compared to derivatives with respect to $y_{i>1}$, an approximation stemming from the slaving assumption [19]. We will specialize the theoretical calculations to our case of $N = 5$ using the feed-forward coupling scheme embodied in (16).

We will be concerned with steady state solutions for the fast and slow element probability density functions. By this we mean that, in the absence of explicit time-dependent terms on the rhs of (19), the solution to the FPE at long times provides a very good representation of the system behavior. In other words, we are concerned with the solution $h(y_2\ldots y_N,t)$ in the $t \to \infty$ limit. In a practical system, this limit is reached after allowing the transient behavior of the system to die down; the rate at which this occurs depends on the time constants a_i, with the fast elements $y_2\ldots y_N$ settling down to their steady states very rapidly, compared to the slow element y_1. Theoretically, the long time limit corresponds to setting the lhs of (19) equal to zero and integrating the resultant equation; this approach is valid only when the rhs of (19) does not contain explicitly time-dependent terms [23].

The utility of the FPE is now clear. As long as one can decouple fast and slow variables leading to the separate FPEs (19) and (20), the steady state solution becomes a matter of integrating an ordinary differential equation subject to the appropriate boundary conditions. This steady state solution which can be (in our system) written down analytically, using a procedure analogous to our derivation of (10), is precisely what we will need in our characterization of the XOR function. Conveniently, the FPEs to be integrated do not involve the noise terms directly, only their moments (in this case the mean and variance); this is because the FPE is a diffusion type equation, for a probability density function, in contrast to the original SDEs (1) which involve the original random variables y_i and, explicitly, the noise terms.

We note that the theoretical approach (underpinned by a steady state solution to the FPE) is equivalent to (numerically) integrating the coupled system (1) and fitting a probability density function to the ensuing time-series solution. As long as we have allowed transients to die out (in practice this can be accomplished by discarding a tranche of data at the front-end of the time series), this numerically generated density function corresponds to the long time solutions to (19) and (20). Clearly, an experimental manifestation of the coupled system, as carried out in Section IV, wherein data is collected as a time series, would yield the same probability density function as the above-described procedure involving numerical integration of (1). Finally, when making comparisons between numerical integration results and nonlinear circuit simulations, a rigorous conversion of abstract dimensionless quantities (e.g. J_{ij}, a_i, σ_i^2) to corresponding quantities (and their correct units) defined via the circuit, is of great importance; we do this in Section IV.

Our choice of coupling scheme allows us to further simplify the system. We note that the dynamics of the input elements $y_{4,5}$ are independent of the remaining elements, due to the absence of back-coupling. Similarly, the elements $y_{2,3}$ can be regarded as independent, having no direct coupling to one another, no back-coupling to the input elements,and only forward coupling to the readout element y_1. For this special case, we can factor the entire density function for the fast variables according to

$$h(y_2, y_3, y_4, y_5) = \prod_{i=2\ldots5} h(y_i) \tag{22}$$

which we substitute into (21) and work systematically through the integrals, remembering that each density function is normalized to unity. As an example, we can write the steady state solutions for the input elements $y_{4,5}$ as follows

$$h(y_4) = \mathcal{N}_4^{-1}\ \exp\left[-\frac{2}{\sigma_4^2}U(y_4)\right]$$
$$U(y_4) = \frac{a_4}{2}y_4^2 - J_{44}\ln\cosh y_4 - \varepsilon_4 y_4 \tag{23}$$

with a similar expression for $h(y_5)$. Here, $\mathcal{N}_4$ is the normalization constant; note that each elemental probability density function is normalized to unity. These solutions are obtained by solving the FPEs for the steady state density functions $h(y_4)$ and $h(y_5)$; in the absence of back-coupling these FPEs do not depend on the elements y_2, y_3) to which (y_4, y_5) are forward coupled only.

Knowing the steady state density functions for the input elements, we can write the dynamics of $y_{2,3}$ in the form:

$$\dot{y}_2 = -a_2y_2 + J_{22}\tanh y_2 + C_2 + N_2(t),$$
$$C_2 \equiv \varepsilon_2 + J_{24}\overline{\tanh y_4} + J_{25}\overline{\tanh y_5} \tag{24}$$
$$\dot{y}_3 = -a_3y_3 + J_{33}\tanh y_3 + C_3 + N_3(t),$$
$$C_3 \equiv \varepsilon_3 + J_{34}\overline{\tanh y_4} + J_{35}\overline{\tanh y_5}. \tag{25}$$

Here we have defined

$$\begin{aligned}\overline{\tanh y_4} &\equiv \int_{-\infty}^{\infty} \tanh y_4\, h(y_4)\, dy_4 \\ \overline{\tanh y_5} &\equiv \int_{-\infty}^{\infty} \tanh y_5\, h(y_5)\, dy_5 . \end{aligned} \tag{26}$$

The effect of the input elements on the intermediate layer (comprising the elements $y_{2,3}$) is to offset the dynamics by constant terms $C_{2,3}$ which adjust the asymmetry in the $y_{2,3}$ dynamics. As noted in the preceding subsection, this results in a skewing of the potential energy functions with a concomitant effect on the corresponding probability density functions (see (7) and (10).

We now address the integrals in (26) using the y_4 integration as an illustration. We assume that the density function $h(y_4)$ sharply peaks at approximately its two stable steady states $y_{4\pm}$. Then one can replace the integral by the sum of two separate integrals in the left and right quadrants, and expand the exponents to second order about each stable state point $y_{4\pm}$ (the first-order terms vanish because the stable states are extrema; in this case, the maxima of the probability density function) and break up each integral into two summed integrals over $[-\infty, 0]$ and $[0, \infty]$. We note the necessity of choosing parameters such that the integrands are sharply peaked about the stable minima of the potential energy function (or the maxima of the corresponding probability density function). The normalization is carried out in the same manner and is included in the calculation. After some work we obtain

$$\overline{\tanh y_4} \simeq \mathscr{A}/\mathscr{B} \tag{27}$$

where

$$\begin{aligned}\mathscr{A} &= \sqrt{\frac{\pi\sigma_4^2}{U''(y_{4-})}} \tanh y_{4-} \exp\left[-\frac{2}{\sigma_4^2}U(y_{4-})\right] \\ &\quad + (y_{4-} \to y_{4+}) \\ \mathscr{B} &= \sqrt{\frac{\pi\sigma_4^2}{U''(y_{4-})}} \exp\left[-\frac{2}{\sigma_4^2}U(y_{4-})\right] \\ &\quad + (y_{4-} \to y_{4+}), \end{aligned} \tag{28}$$

where the double prime denotes the second derivative, and the steady states are given by (see the preceding subsection)

$$\begin{aligned} y_{4\pm} &= \pm y_{40} + \Delta_4 \\ \Delta_4 &\equiv \varepsilon_4/(a_4 - J_{44}\,\mathrm{sech}^2\, y_{40}) \\ y_{40} &= \frac{J_{44}}{a_4} \tanh \frac{J_{44}}{a_4}. \end{aligned} \tag{29}$$

We can do some additional simplification to finally obtain

$$\overline{\tanh y_4} \simeq \frac{\tanh y_{4-} + f_4 \tanh y_{4+} \exp\left[-\frac{2}{\sigma_4^2}\Delta U_4\right]}{1 + f_4 \exp\left[-\frac{2}{\sigma_4^2}\Delta U_4\right]} \tag{30}$$

with the definitions

$$f_4 \equiv \sqrt{\frac{a_4 - J_{44}\,\mathrm{sech}^2\, y_{4-}}{a_4 - J_{44}\,\mathrm{sech}^2\, y_{4+}}}, \quad \Delta U_4 \equiv U_4(y_{4+}) - U_4(y_{4-}). \tag{31}$$

In (27) and (30), the denominators arise via the normalizations. An analogous (to (30)) expression can be written for $\overline{\tanh y_5}$. Thus, we have evaluated the constants $C_{2,3}$ on the rhs of (25); the input signals $\varepsilon_{4,5}$ are contained in these constant terms, together with all other parameters (including the noise variances) contained in the input elements.

Equations (30) and (31) can be simplified further. We can easily set $f_4 \approx 1$ because of the behavior of $\mathrm{sech}\, x$ at large x; this assumes that the parameters (a_4, J_{44}) are selected such that $U(y_4)$ is bistable and the (stable) fixed points are reasonably well separated from the unstable fixed point. Further, we can calculate the quantity ΔU_4

$$\begin{aligned}\Delta U_4 &= U(y_{4+}) - U(y_{4-}) \\ &\equiv U(y_{40} + \Delta_4) - U(-y_{40} + \Delta_4). \end{aligned} \tag{32}$$

Using the definition in (29), we are readily led to

$$\Delta U_4 \simeq -2\varepsilon_4 y_{40} \tag{33}$$

with a similar result for ΔU_5. To arrive at (33), we perform a Taylor expansion of (32) to $O(\Delta_4)$. We now consider (30), and use the definitions in (29) for the fixed points, to expand the hyperbolic functions to $O(\Delta_4)$. After the calculation we obtain

$$\overline{\tanh y_4} \approx \tanh\left(\frac{2y_{40}\varepsilon_4}{\sigma_4^2}\right) + \Delta_4\,\mathrm{sech}^2\, y_{40} \tag{34}$$

where we have set $\tanh y_{40} \approx 1$. Following the same line of calculation

$$\overline{\tanh y_5} \approx \tanh\left(\frac{2y_{50}\varepsilon_5}{\sigma_5^2}\right) + \Delta_5\,\mathrm{sech}^2\, y_{50}. \tag{35}$$

With these evaluations, the equations for $y_{2,3}$ are closed and we can write the steady state density functions

$$\begin{aligned} h(y_2) &= \mathscr{N}_2^{-1} \exp\left[-\frac{2}{\sigma_2^2}U(y_2)\right], \\ U(y_2) &= \frac{a_2}{2}y_2^2 - J_{22}\ln\cosh y_2 - C_2 y_2 \end{aligned} \tag{36}$$

with an analogous expression for $h(y_3)$.

Let us, now, recall that the y_1 dynamics are given by

$$\dot{y}_1 = A(y_1) + N_1(t), \tag{37}$$

with the definitions

$$\begin{aligned} A(y_1) &= -a_1 y_1 + J_{11}\tanh y_1 + C_1 \\ C_1 &\equiv \varepsilon_1 + J_{12}\overline{\tanh y_2} + J_{13}\overline{\tanh y_3} \end{aligned} \tag{38}$$

and, as before,

$$\overline{\tanh y_2} \equiv \int_{-\infty}^{\infty} \tanh y_2 h_2(y_2) dy_2$$
$$\overline{\tanh y_3} \equiv \int_{-\infty}^{\infty} \tanh y_3 h_3(y_3) dy_3. \tag{39}$$

With the (now) closed form expression (36) and the analogous expression for $h(y_3)$, we can compute the integrals above exactly as done previously. Finally, we are led to the steady state probability density function (10) for the reference (or readout) element y_1 in closed form, with the potential energy function

$$U(y_1) = \frac{a_1}{2} y_1^2 - J_{11} \ln\cosh y_1 - C_1 y_1 \tag{40}$$

which has its extrema located at

$$\begin{aligned} y_{1\pm} &= \pm y_{10} + \Delta_1 \\ \Delta_1 &\equiv C_1/(a_1 - J_{11}\,\mathrm{sech}^2\, y_{10}) \\ y_{10} &= \frac{J_{11}}{a_1} \tanh \frac{J_{11}}{a_1} \end{aligned} \tag{41}$$

with $\mathcal{N}_1$ being the normalization.

The above results led us to consolidate the all-important expressions for C_i

$$\begin{aligned} C_2 &= \varepsilon_2 + J_{24} \tanh\left(\frac{2y_{40}\varepsilon_4}{\sigma_4^2}\right) + J_{25} \tanh\left(\frac{2y_{50}\varepsilon_5}{\sigma_5^2}\right) \\ C_3 &= \varepsilon_3 + J_{34} \tanh\left(\frac{2y_{40}\varepsilon_4}{\sigma_4^2}\right) + J_{35} \tanh\left(\frac{2y_{50}\varepsilon_5}{\sigma_5^2}\right) \\ C_1 &= \varepsilon_1 + J_{12} \tanh\left(\frac{2y_{20}C_2}{\sigma_2^2}\right) + J_{13} \tanh\left(\frac{2y_{30}C_3}{\sigma_3^2}\right) \end{aligned} \tag{42}$$

where we have dropped the second terms on the rhs of (34) and (35), as well as in the corresponding (not shown) calculated quantities from (39), because (for all reasonable parameters) these terms are quite small. Importantly, the expansion in e.g. (29) and (41) require the second (and higher order) terms to be $\ll 1$. It is possible that these terms are ≈ 1, depending on the choice of parameters. However, multiplying these terms by $\mathrm{sech}^2\, y_{i0}, i = 2 \ldots 5$ (and higher powers thereof) renders the second (and higher) terms on the rhs of e.g. (34), (35) $\ll 1$ so that they can be dropped when compared to the leading terms.

Equation (40), in which the *deterministic* constant C_1 contains the net effects of coupling to the rest of the network, is used to realize the XOR gate based on the structure of $U(y_1)$ or, equivalently, $P(y_1)$. We note that the effect of the fast elements is to contribute to the asymmetry (corresponding to $C_1 \neq 0$) of the probability density function $P(y_1)$ of the reference (or readout) element. The noise variances in the input elements and the intermediate layer are also incorporated in C_1, together with all the other system parameters appearing in (16). Finally we observe that, for the special case of all $\varepsilon_i = 0$, the integrals in (26) and (39) vanish such that the potential energy function $U(y_1)$ and the associated probability density function $P(y_1)$ are bistable and symmetric.

Later (Section IV), we describe circuit elements that can replicate this system in a CMOS environment. For now, however, we close this section with an important note. Our system (Figure 1) does not have back-coupling, meaning that terms in J_{21}, J_{31} etc., are taken to be zero. In turn this allows a decomposition of the "bath" density function into the product of individual density functions, as given in (22). For our particular system (1) without back-coupling, this means that the fundamental tenet of the adiabatic (or slaving) assumption ($a_1 \ll a_{i>1}$) can be relaxed. However, in the presence of back-coupling, this assumption becomes necessary to achieve a smooth decoupling of the system into slow and fast elements; this is exemplified in our earlier work [12]. With non-zero back-coupling, the calculation is more complicated, however, it remains tractable. Regardless, we chose to display, formally in (17)-(21), the decoupling via the "slaving principle", to demonstrate how it could be carried out for the more general case. It should be clear that our procedure (even with back-coupling included) can be applied to larger arrays even though, in this work, we focus on a small feed-forward network. In our earlier work [12] we did, in fact, treat a simplified network with back coupling, using the slaving principle to achieve the decoupling of fast and slow elements.

Before concluding this subsection, we return to the consideration of the minimum number of elements required to realize the XOR in our dynamical network. Focusing on the sub-network defined by the elements 3, 4 and 5, and viewing it as an isolated single-layer network with two inputs and one output (or readout) element, the expression for C_3 in (42) provides a mapping from the input variables to the effective offset or skewing constant which completely characterizes the probability density of the output element:

$$\{\varepsilon_4, \varepsilon_5\} \longmapsto \varepsilon_3 + \sum_{i=4}^{5} \alpha_i \tanh \beta_i \varepsilon_i \tag{43}$$

Here, we have subsumed the constants appearing in (42) into α_i and β_i. After suitable re-indexing and re-labeling, this functional form should be compared with (15) and is readily seen to be manifestly inadequate to realize the XOR. This stems from the monotonicity of the tanh function and the separation of the input variables in the arguments. Therefore, in order to realize the XOR with our dynamical network, we need a hidden layer containing at least two elements, so that the total number of dynamical elements is $N = 5$; this is similar to the case with conventional ANNs. We note that for the case of ANNs, the requirement that $N = 5$ at minimum to realize the XOR is described in [21] and [22].

C. STRUCTURE OF THE PROBABILITY DENSITY FUNCTION $P(Y_1)$

The preceding analysis has shown that the effects of the background noise terms in the fast elements enter into the

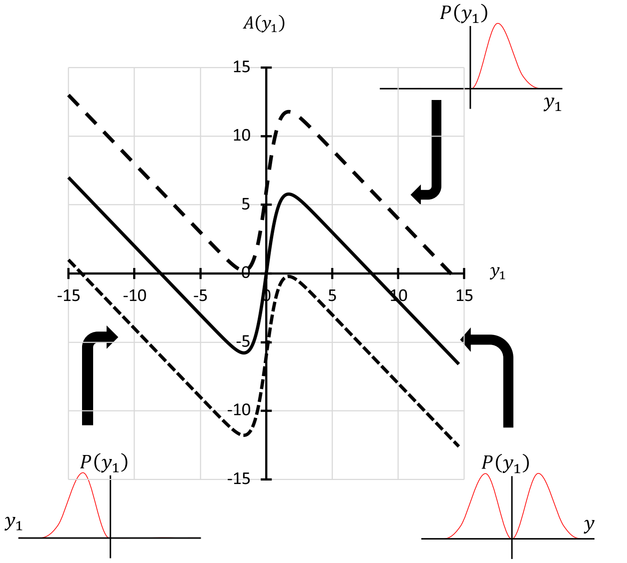


FIGURE 2. Transition to monostability. The dashed curves depict the transfer characteristic $A(y_1)$ corresponding to the critical cases (inequality replaced by equality) of equations (46) (bottom) and (47) (top). Middle (solid curve) corresponds to $C_1 = 0$ (symmetric bistability). The critical points (switching curve touching the horizontal axis) are y_p (bottom curve) and y_m (top curve).

slow element dynamics via the constant term C_1 defined in (38). The effect of this constant term is to skew the density function $P(y_1)$. If $C_1 = 0$ (which can be achieved by judicious adjustment of the magnitude and sign of the dc bias term ε_1, or by adjusting other parameters in (16)), the density function is bimodal and symmetric, with the peaks centered about the *deterministic* fixed points $y_{1\pm}$. The transition between mono- and bi-modality is governed by the streaming term $A(y_1)$ (38), whose zeros yield the locations of the extrema of the potential energy function (this point was also discussed in our analysis of a single bistable element, above). The extrema of the probability density function $P(y_1)$ are, exactly, the extrema of the potential energy function $U(y_1)$ with the minima of the potential energy function corresponding to the maxima in the probability density, and the unstable fixed point between the potential minima corresponding to the minimum (located between the 2 maxima) of the probability density.

The transition from bistability to monostability corresponds to the three real roots of the transfer function $A(y_1)$ collapsing into a single real root and two complex conjugate roots. It is worth quantifying the threshold at which this occurs. Figure 2 shows a plot of the function $A(y_1)$ for an arbitrary (for purposes of illustration) choice of parameters a_1, J_{11} selected so that, in the absence of a constant term on the rhs, the system is bistable, meaning that $A(y_1)$ has three real roots; this means taking $|J_{11}| > |a_1|$. In the middle curve we set $C_1 = 0$ yielding a transfer characteristic that is symmetric about the vertical axis and has three real roots. The potential energy function for this case is bistable with wells of equal depth with the minima occurring at the intersections of $A(y_1)$ with the horizontal axis. These minima are also, as mentioned above, the locations of the peaks in the probability density function obtained by using (40). The unstable point of the potential is at $y_1 = 0$.

As C_1 steadily decreases from zero, the area between the curve and the positive axis in the right half plain steadily decreases. This corresponds to one of the potential wells becoming deeper than the other until, at the critical value (shown in the bottom curve), a minimum in the potential energy function (in this case, on the right side) has been replaced by an inflection point (occurring at $y = y_p$). Past this point the transfer function admits one real root and two complex conjugate roots; the potential is monostable (but not parabolic). Completely analogous changes occur in the steady state probability density function; it transitions from bimodal (and symmetric) to unimodal (in this illustration, on the left side, with the peak located at the real root of $A(y_1)$.

Simple calculus allows us to quantify this transition. We begin with the streaming term in Equation (38). At $y_1 = y_p$ (the bottom curve in Figure 2), one has an extremum so that $A'(y_p) = 0$. This leads to $y_p = \tanh^{-1}\sqrt{1 - \frac{a_1}{J_{11}}}$. Very close to the critical case (for the bottom curve in Figure 2), we can expand $A(y_1)$ about y_p

$$A(y_1) = A(y_p) - a_1(y_1 - y_p)^2 \tanh y_p. \tag{44}$$

Rearranging terms, we obtain a quadratic in y_1 which can be solved for the roots

$$y_1 = y_p \pm \left[\frac{A(y_p)}{a_1 \tanh y_p}\right]^{1/2} \tag{45}$$

where $\tanh y_p$ is given above. Referring to Figure 2 it is clear that, for the transition occurring at $y_1 = y_p$ we must have $A(y_p) > 0$ for bistability (3 real roots), $A(y_p) = 0$ at the critical point (where the bottom curve just touches the y_1 axis), and $A(y_p) < 0$ for one real and two complex conjugate roots; in the last case, the single real (and negative) intersection of the transfer curve with the y_1 axis marks the location of the (single) peak of the monomodal density function. Putting all this together, we obtain

$$A(y_p) = -a_1 \tanh^{-1}\sqrt{1 - \frac{a_1}{J_{11}}} + J_{11}\sqrt{1 - \frac{a_1}{J_{11}}} + C_1 < 0 \tag{46}$$

as the condition for monostability in the Left Half Plane (LHP). The critical case (shown in the bottom curve of Figure 2) occurs for $A(y_p) = 0$. In an entirely analogous manner (in this case, by increasing C_1) we can obtain a monomodal density function for the Right Half Plane (RHP). The condition for this is found to be

$$A(y_m) = a_1 \tanh^{-1}\sqrt{1 - \frac{a_1}{J_{11}}} - J_{11}\sqrt{1 - \frac{a_1}{J_{11}}} + C_1 > 0 \tag{47}$$

with the critical case now occurring for $A(y_m) = 0$. The rhs of the above equalities determine the critical values of the constant term C_1 as the density function transitions from being monostable (in the LHP) and peaked at the positive root of the transfer characteristic (bottom curve in (2)), to being monostable in the RHP (top curve). Between these extreme cases, various "degrees" of bistability are obtained. We note that the mean value $\langle y_1 \rangle = \int_{-\infty}^{\infty} y_1 P(y_1) dy_1$ traverses from right to left as we progress through the situations (top-to-bottom) depicted in Figure 2. Concomitantly, the mean value $\langle y_1 \rangle = \int_{-\infty}^{\infty} y_1 P(y_1) dy_1$ traverses from left to right as we progress through the situations (top-to-bottom) depicted in Figure 2. When $P(y_1)$ is perfectly monomodal the mean value corresponds to the location (given by Equation (41)) of the peak. For the middle case in Figure 2, the mean value is zero, corresponding to $C_1 = 0$ in (38), that is, a symmetric bimodal probability density function $P(y_1)$.

Clearly, with all other parameters fixed, the constant forcing term ε_1 may be adjusted to realize the critical (from (46) or (47)), or any other desired value (from (42)) of C_1. If ε_1 is to be kept fixed (or zero), the other system parameters in (16) ca be adjusted to realize the required value of C_1. This is a fairly complicated proposition because of the sheer number of available parameters. We reiterate that, in general, the crossing points of the switching curve $A(y_1) = 0$, as shown in Figure 2, are functions of every element in the network i.e. they depend on all the parameters in the array (16) as well as the noise variances $\sigma_{i>1}^2$ of the "fast" or "bath" elements.

The above analysis, aimed at determining the characteristics of the probability density function $P(y_1)$, e.g. the locations and relative heights of the peaks, are essentially *deterministic*, containing the noise variances $\sigma_{i>1}^2$ as parameters that change the peak locations and heights through their effect on the quantities (26) and (39) and, ultimately, on $A(y_1)$, the deterministic contribution to the y_1 dynamics (37). The slow element noise variance σ_1^2 appears only in the slow dynamics (20) leading to the long time behavior quantified in (40); its sole effect is to change the width of the lobes of the probability density function $P(y_1)$. It does *not* affect the locations of the peaks of $P(y_1)$.

We can investigate this behavior further by examining the regimes where the constant C_1 forces the probability density $P(y_1)$ to transition from bi- to mono-modality; this is central to our characterization of the XOR and is done in the following sections.

III. APPLICATION OF THE THEORY TO THE XOR TRANSFORM

In what follows, we apply the theoretical results of Section II to the realization of the analog equivalent of the XOR function in probability space. We show that this network provides a realization of an XOR (predicated on the structure of the probability density function $P(y_1)$ of the readout element) except in well-defined regions of parameter space. In a subsequent section, we validate our theoretical results with numerical simulations and propose a nonlinear circuit implementation of the 5-element system.

We begin by recapping the observation that the probability density $P(y_1)$ can transition between mono- and bi-modality as shown in Figure 2. The transition between these configurations is controlled by the parameter C_1 as calculated in (38) and (42). For $C_1 = 0$ the probability density function is bimodal with symmetrically placed peaks of equal heights. As C_1 starts to deviate from zero, the density function transitions towards monostability in the left or right half, depending on the sign of C_1. The critical point at which purely monostable behavior is obtained can be calculated using (46) or (47). Given the large number of parameters in our 5-element system, we keep the matrix of couplings J_{ij} fixed in our analysis going forward, while ensuring that all elements are bistable, that is $J_{ii}/a_i > 1$. We then examine the XOR in probability space by varying only the noise variances σ_i^2 and the constant bias terms ε_i.

Our approach consists of calculating the parameter C_1 as a function of $\varepsilon_{4,5}$, the constant biases applied to the input layer. As a representative case, Figure 3 shows the contours of C_1 in the 2D space of $\varepsilon_{4,5}$. The long time probability density function $P(y_1)$ is symmetrically bimodal along the contour $C_1 = 0$ shown in the figure. On either side of the $C_1 = 0$ contours, we have a gradual transition to monomodality with the critical C_1 calculated using (46) or (47). For this specific set of parameters, the $C_1 = 0$ contours serve as delineations between the two logic output states. Different bias bias and system parameters can create different (but still bistable) probability distributions than those in Figure 3, but the requirement is the same: two different distinguishable distributions, one resulting from $(\varepsilon_4, \varepsilon_5)$ in quadrants I and III, the other from $(\varepsilon_4, \varepsilon_5)$ in quadrants II and IV.

Note that, due to the immense number of adjustable parameters in the theoretical model, we have selected a particular set of parameters, for the simulations. In particular, we keep the matrix J_{ik}, $i, k = 1 \ldots 5$, the column vector of time constants a_i, $i = 1 \ldots 5$, the dc biases $\varepsilon_{1,2,3}$, and the noise variances σ_i^2 $i = 1 \ldots 5$ fixed. This allows us to examine the XOR in the space of the two input dc bias signals $\varepsilon_{4,5}$. These parameters are kept fixed for the numerical simulations, as well. It is clear, that changing the magnitudes and signs of some of the parameters, e.g. the coupling coefficients, could lead to much different behavior, including changing where the XOR is realized in the space of $\varepsilon_{4,5}$. For this paper, we deliberately chose parameters such that every element in the system (1) remains bistable, assuming that the bias signals are chosen to be small enough to retain the bistability in each element.

The ratios of the peak heights in the density function $P(y_1)$ are important for characterizing the XOR; this is evident from Figure 3. These can be calculated in terms of the density function $P(y_1)$ or its associated potential energy function $U(y_1)$. We begin with the potential energy function. The difference (underpinned by the constant C_1) in well-depths can be expressed as

$$\Delta U_1 \equiv U(y_{1-}) - U(y_{1+}) = 2C_1 y_{10} \tag{48}$$

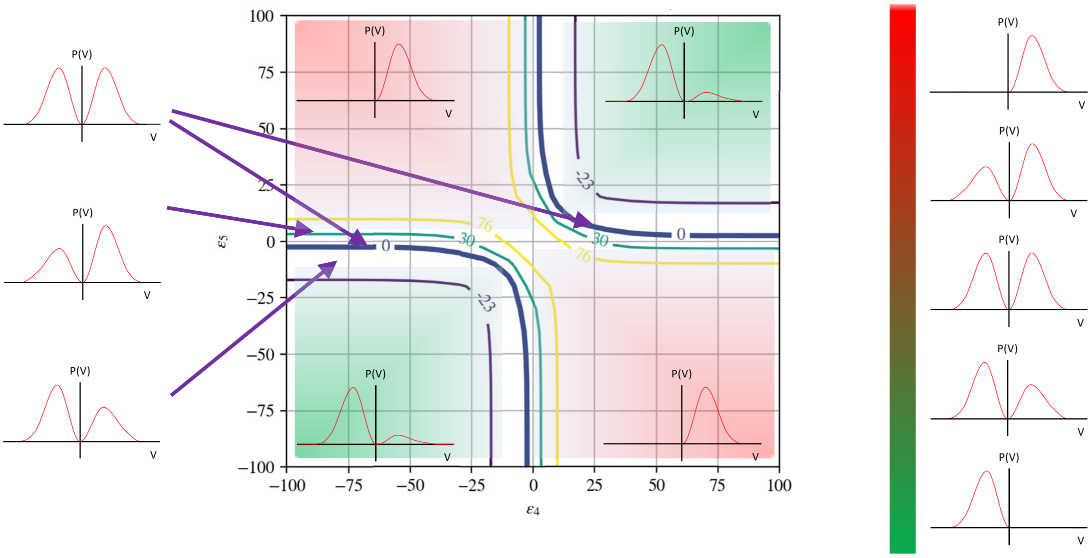


FIGURE 3. Topographical representation defining an XOR transform as a function of the constant inputs ε_4 and ε_5. Green region (inputs $\varepsilon_{4,5}$ have the same signs) represents a probability distribution favoring a negative value of the random variable y_1 (nominally "FALSE"), whereas red ($\varepsilon_{4,5}$ have opposite signs) represents a probability distribution favoring a positive y_1 (nominally "TRUE"). Other parameters are: $J_{11} = 100, J_{ii} = 500$ $(i > 1), J_{12} = J_{13} = J_{24} = J_{25} = 60, J_{34} = J_{35} = -60, a_1 = 10, a_i = 100$ $(i > 1), \varepsilon_1 = -25, \varepsilon_{2,3} = 50, \sigma_i^2 = 100$ $(i = 1 \ldots 5)$.

where we have used (40) and (41) and expanded to leading order in Δ_1. Given that $y_{10} = (J_{11}/a_1)\tanh(J_{11}/a_1)$ is positive by definition, the sign of C_1 determines which well is deeper. A similar calculation can be carried out for the *ratio* of the peak heights in the probability density function $P(y_1)$

$$\Delta P_1 \equiv P(y_{1-})/P(y_{1+}) = \exp(-2C_1 y_{10}) \tag{49}$$

which is also, approximately, the ratio of the areas enclosed by each peak when using the Gaussian approximation (as was performed in Section II). This ratio is the ratio of the probabilities of the system being in one state (negative quadrant) or the other (positive quadrant). In the event of equal probability of being in the two states (that is, perfectly symmetric bistability), $C_1 = 0$ and the ratio is unity, as expected.

Analogous to Figure 3, we can examine the structure of the probability density function $P(y_1)$ as a function of the inputs $\varepsilon_4, \varepsilon_5$ for different selections of the input noise. This is illustrated in Figure 4. In each sub-figure the case for $C_1 = 0$ (corresponding to symmetric bistability) is shown. We recall that the noise variance in the elements y_k, where $k = 2, 3, 4, 5$ can affect the structure (including the fixed points) of the individual density functions $h(y_k)$ and, hence, the structure of the readout element density function $P(y_1)$; this is evident when we compare the figures. As an example, in the bottom center figure where noise is increased in both the hidden and input layers, the transition to complete monostability in the RHP occurs for $C_1 = 76$, within the red regime depicted in Figure 3. The critical points for transition to monostability can be calculated using Equations (46) and (47). We reiterate that the noise variance σ_1^2 in the readout element only affects the spread of its density function, with no effects on the fixed points. Accordingly, in Figure 4 we hold σ_1^2 fixed. In addition, for simplicity, the matrix of coupling coefficients J_{ik} and the constants a_i are also kept fixed.

A. COUPLED ELEMENT SIMULATIONS

To validate the theoretical model developed in the preceding section, we solved system (16) numerically. The general stochastic differential equation

$$\begin{aligned} \frac{dy}{dt} &= f(y,t) + \sigma\xi(t) \\ y(0) &= y_0 \end{aligned} \tag{50}$$

is solved by Euler-Maruyama integration [25]. Using the definition of white noise $\xi(t) = dW/dt$ as the derivative of a Wiener process, we rewrite (50) as

$$dy = f(y,t)dt + \sigma dW. \tag{51}$$

Let Δt be the integration step, and let $t_i = i\Delta t$. At t_{i+1}, y is incremented by approximately

$$\Delta y = y(t_{i+1}) - y(t_i) \simeq f(y(t_i), t_i)\Delta t + \sigma \Delta W_i \tag{52}$$

where ΔW_i is drawn from a normal distribution $N(0, \sqrt{\Delta t})$ with mean zero and variance Δt.

The characteristic time scale of the system (1) is defined by the fastest element: $\tau = 1/\max\{a_i\}$. In order for the Euler method to be stable, we must have $\Delta t \ll 2\tau$. Balancing the need for accurate simulations with the execution time, we chose a time step $\Delta t = \tau/B$, and solved (52) for $t \in [0, M\tau]$, thereby generating 5 sequences of $B \cdot M$ points $\{y_i(t_j)\}$. We discard the first few points to let transients

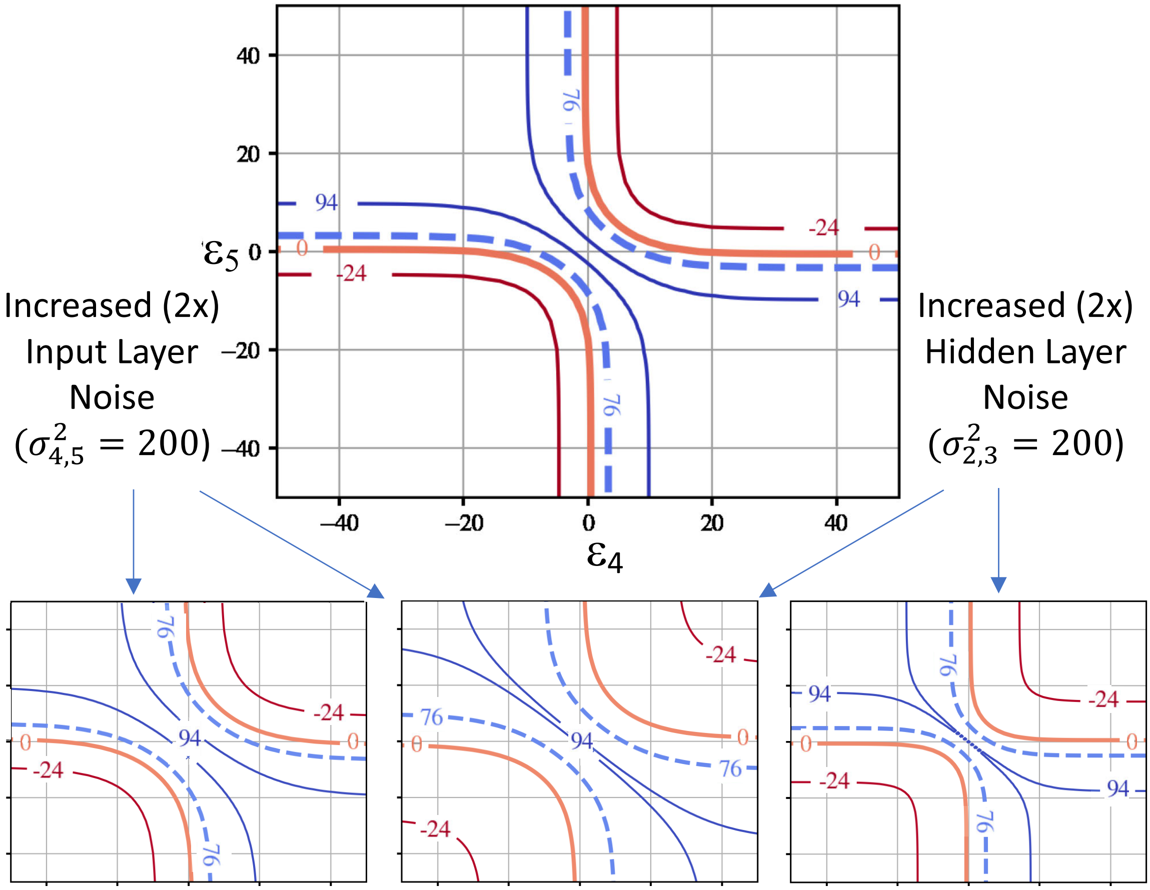


FIGURE 4. Four contour plots of C_1 (as in Figure 3) on the same scale showing the effect of changing noise intensity in the first two layers. The "baseline" case (top panel) has noise variances $\sigma_i^2 = 100$ for all elements i. Contours labeled "0" are the locus of symmetric bistability, while the dashed contours mark the transition to "complete" monostable density functions. Other parameters as in Figure 3. See text for further details.

die down. Then, the steady state probability densities for y_i are calculated by applying a kernel density estimator with bandwidth 0.2 to the resulting sequences, with smoother appearing densities for larger values of M. For Figure 5, we used $B = 10$ and $M = 10^6$.

In Figure 5 we plot the simulated steady-state probability density functions for all elements y_i in response to four pairs of Boolean inputs defined by pairs $(\varepsilon_4, \varepsilon_5)$ where $\varepsilon_{4,5} = \pm 7$ (all other ε_i are fixed). When the two inputs have the same signs (i.e. either both at -7, or both at +7, corresponding to ("FALSE", "FALSE") or ("TRUE", "TRUE")) the distribution $P(y_1)$ is left-skewed and an output of "FALSE" is expected. However, when the system is presented with inputs having opposite signs (i.e. one at -7, the other at $+7$, corresponding to ("FALSE", "TRUE")), the distribution $P(y_1)$ is right-skewed and the output "TRUE" is expected. In addition, the corresponding theoretical steady-state (i.e. in the $t \to \infty$ limit) probability density functions (23), (36), and (40), obtained in Section II, demonstrate good agreement (in particular the peak locations and relative heights) with the numerical simulations. While this result is for J_{ij} and (fixed) ε_i that were not selected via any definitive criteria, the critical point is that J and ε are chosen such that the four input pairs reduce to two output probability distributions which are sufficiently distinct that they can be differentiated and equated to representations of logical "TRUE" or "FALSE", thus comprising an XOR transform. The characteristics of the distributions generated from a given choices of J and ε (peak locations, height, etc.) are immaterial as long as a correspondence can be made.

IV. NONLINEAR COUPLED CIRCUIT IMPLEMENTATION

Finally, we implemented the computational method presented here in physically realizable hardware. As fabrication of components is extremely expensive, we chose a technology, CMOS, that has well established simulation suites such as Cadence. Although it is difficult to verify noise driven switching in the Cadence iterative simulation environment, CMOS represents a reasonable first exploration into circuit realizations. It provides initial evidence that our underlying assumptions are valid while offering an environment in which implementation can be validated. Established platforms for integrated circuit design and simulation, such as Cadence, are not designed for validating transient solutions in systems wherein the noise to signal environments approach unity. As a result Cadence is slow and difficult if not impossible for system validation in the time domain. We used Cadence to simulate the coupled $N = 5$ system without added

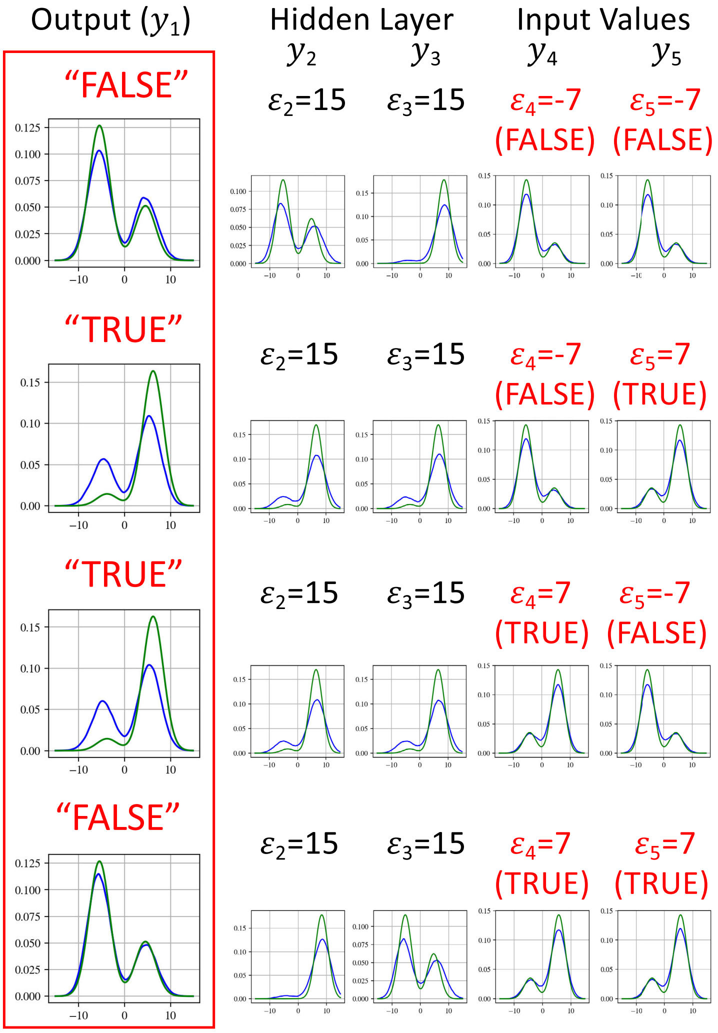


FIGURE 5. XOR demonstration via simulated steady state probability density function $P(y_1)$ of readout element. In this example, the self-coupling terms are $J_{11} = 5$, $J_{ii} = 50$ for $i > 1$. The cross-coupling terms are $J_{12} = J_{13} = 1.5$, $J_{24} = J_{25} = 15$, $J_{34} = J_{35} = -15$. $\varepsilon_1 = -1.5$ Each box shows probability density function derived from theory (Section II) in blue and simulation in green. The theory, despite invoking an expansion of the form (29) to calculate the fixed points, predicts the extrema remarkably well. The theoretical probability density function is calculated in the $t \to \infty$ limit and would yield better overall agreement with the simulations if the latter were run for longer times. Nonetheless, the ratios of peak heights (or equivalently, the ratios of areas enclosed by each peak) show good agreement between theory, i.e. (49), and simulations.

noise (meaning there may still be component noise present), and showed good agreement with the theory of Section II. We demonstrate this comparison of theoretical results to those obtained via the circuit simulation, by observing the hysteresis behavior and comparing it to the (deterministic) switching curves depicted in Figure 2.

A. THE TANH FUNCTION GENERATING CIRCUIT

To construct a physical circuit implementing the dynamical system (16), each non-zero coupling coefficient J_{ij} is embodied by a tanh generating circuit. It is well known that CMOS devices operating in subthreshold are readily configured to achieve a five-device tanh function [26] (see Figure 8 for a simplified depiction of three such circuits implementing row 1 of the J matrix).

A symbolic representation of a prototypical tanh generating circuit $\mathcal{C}$ is shown in red in Figure 6. This circuit produces a differential output current proportional to the tanh of the differential input voltage

$$I_{\rm op} - I_{\rm om} = I \tanh\left(\frac{V_{\rm id}}{2\kappa U_T}\right) \tag{53}$$

where $V_{\rm id} = V_{\rm ip} - V_{\rm im}$ is the differential input voltage, κ is the subthreshold slope (extracted as 1.07 from Global Foundries 12LP), and $U_T = k_B T/q$ is the thermal voltage (where k_B is Boltzmann's constant, T is the absolute temperature, and

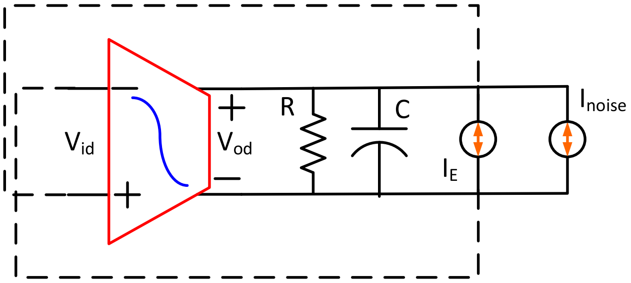


FIGURE 6. Prototypical circuit $\mathcal{C}_{ij}$ with tanh generating block in red. When $i = j$, the dotted lines are connections and $V_{od} = V_{id}$, representing the self-coupling coefficient J_{ii}.

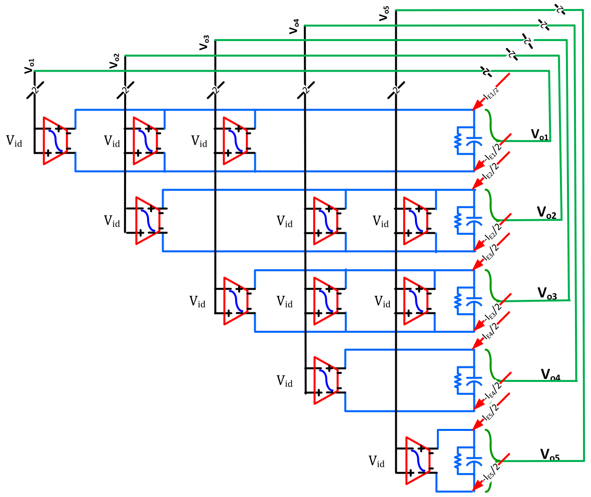


FIGURE 7. Circuit for 5 element realization. Negative weights J_{ij} are implemented by crossing the output wires. Red elements are as displayed in Figure 6.

q is the electronic charge). The differential output voltage is defined to be $V_{\mathrm{od}} = V_{\mathrm{op}} - V_{\mathrm{om}}$.

When $\mathcal{C}$ is connected in parallel with a load resistor R, capacitor C, and external current source I_E as shown in Figure 6, Kirchoff's current law dictates that the state of the circuit evolves according to

$$C\frac{dV_{\mathrm{od}}}{dt} = -\frac{V_{\mathrm{od}}}{R} + I\tanh(bV_{\mathrm{id}}) + I_E + I_{\mathrm{noise}} \tag{54}$$

where $b = 1/(2\kappa U_T)$. The noise current I_{noise} is explicitly defined later in this paper. Here, C includes both the intrinsic and extrinsic parasitics of the CMOS cell.

For each non-zero coupling coefficient J_{ij}, we create a fully differential circuit $\mathcal{C}_{ij}$ with tail current I_{ij} (explicitly defined later), input terminals at voltages $V_{\mathrm{ip}j}$, $V_{\mathrm{im}j}$ and output terminals at voltages $V_{\mathrm{op}i}$, $V_{\mathrm{om}i}$. All $\mathcal{C}_{i*}$ outputs in row i are at the same voltage as the input terminals of $\mathcal{C}_{ii}$. All $\mathcal{C}_{*j}$ inputs in column j are at the same voltage as the output terminals of $\mathcal{C}_{jj}$. Figure 7 shows a block level circuit topology for the network of five elements implementing the XOR function.

To complete the mapping of this network to a system of stochastic differential equations, we define the state of the i-th element as the differential voltage of $\mathcal{C}_{ii}$

$$v_i = V_{\mathrm{id}i} = V_{\mathrm{od}i} = V_{\mathrm{op}i} - V_{\mathrm{om}i}. \tag{55}$$

Inserting indices into (54) as appropriate, v_i obeys the state equation

$$C_i\frac{dv_i}{dt} = -\frac{v_i}{R_i} + \sum_j I_{ij}\tanh(bv_j) + I_{E_i} + s_i\xi(t). \tag{56}$$

The noise current $I_{\mathrm{noise}i}$ is modeled as an additive Gaussian white noise signal $N_i(t) = s_i\xi(t)$ of average power spectral density [26]

$$s_i^2 = \frac{2}{3}4\,k_BT\frac{I_{\mathrm{tran}i}}{2\kappa U_T} = \frac{4q}{3\kappa}I_{\mathrm{tran}i} \tag{57}$$

where $I_{\mathrm{tran}i} = 2\sum_j I_{ij}$ is the current through the transistors in row i which contribute to the noise, and q is the charge of an electron.

By substituting $y_i = bv_i$ and dividing by C_i/b, we obtain the equations as described in Section II:

$$\frac{dy_i}{dt} = -a_iy_i + \sum_j J_{ij}\tanh y_j + \varepsilon_i + \sigma_i\xi(t). \tag{58}$$

The conversions between circuit parameters and abstract dynamical system are: $a_i = (R_iC_i)^{-1}$, $J_{ij} = bI_{ij}/C_i$, $\varepsilon_i = bI_{E_i}/C_i$, $\sigma_i = bs_i/C_i$.

Figure 7 introduces a block level circuit topology that implements the set of equations in (16). It uses the following parameters:

$$\begin{aligned}
&I_{11} = 105, I_{jj} = 500(j > 1), I_{12} = 25 = I_{13},\\
&I_{24} = 100 = I_{25}, I_{34} = -100 = I_{35}(nA)\\
&R_1 = 1.4, R_{k>1} = 0.4(\mathrm{M}\Omega); C_{11} = 500, C_{k>1} = 100(\mathrm{fF})\\
&I_{E_1} = 0, I_{E_2} = I_{E_3} = 50(\mathrm{nA})\\
&\epsilon_4 = \pm 50, \epsilon_5 = \pm 50(\mathrm{nA}) \text{ are “TRUE”/“FALSE”}\\
&\qquad \text{network inputs.}
\end{aligned}$$

Noise computations for the circuit in Figure 7 can be derived from the methods presented in [26]. The effective bandwidth of the idealized white noise signal $N_i(t)$ is approximately $\Omega_i = 1/(4R_iC_i)$, resulting in a noise current with 1-sigma amplitude

$$s_i\sqrt{\Omega_i} = \sqrt{\frac{q}{3\kappa}\frac{I_{\mathrm{tran}i}}{R_iC_i}} \simeq 0.1\text{ nA} \tag{59}$$

which is significantly less than the typical noise amplitudes used in the simulations of (58). Based on these estimated results we expect to have to inject noise into each diagonal element of Figure 7. The signal-to-noise ratio desired is a trade off between acquiring a satisfactory sample size and the bandwidth of the inputs.

A simplified schematic of the implementation of element v_1 is shown in Figure 8. Note that Figure 8 is the basis for all five rows in Figure 7. Each term $I_{ij}\tanh(bv_{ij}(t))$ of Equation (56) is implemented by MOSFET devices M1-M5, with the tail currents of M5 setting the weights I_{ij}. In this example, the current through M5A sets the diagonal weight I_{11}, whereas the currents through M5B and M5C set the off-diagonal weights I_{12}, I_{13}. Cascode transistors M3 and M4

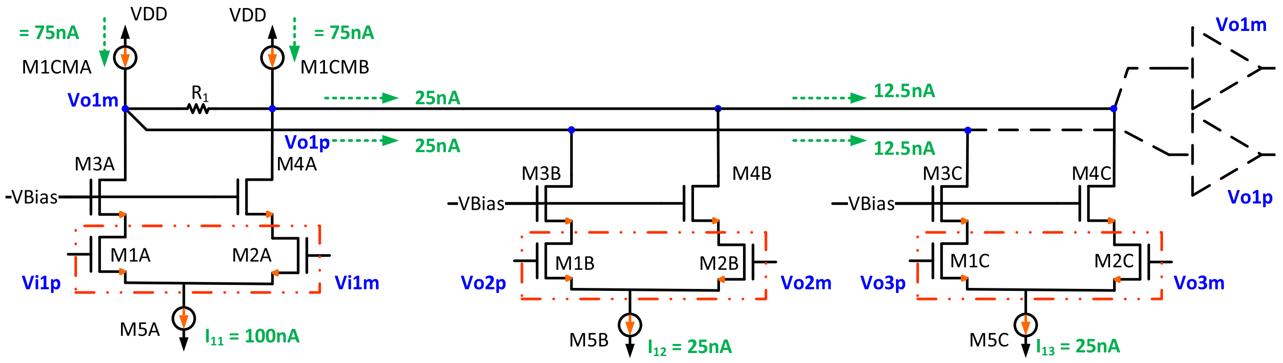


FIGURE 8. Simplified circuit representing a general element of layer 1 and 2. The left most cell is the diagonal tanh element providing self-coupling, and the right two are the off-diagonal elements providing coupling from the previous layer. Pin and current labels within the figure are specific to element v_1. Green arrows at the top represent current flow when all biases $I_{E_i} = 0$.

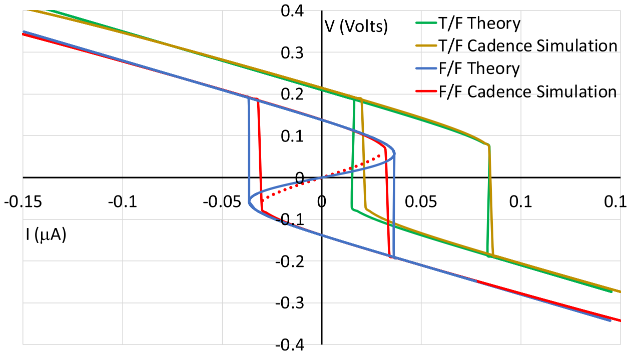


FIGURE 9. Hysteresis loops obtained from Cadence circuit simulation (red/brown) and numerical integration of (56) (blue/green). The XOR="TRUE" case corresponds to the loops on the right (green/brown) which have only one crossing of the vertical axis (monostable probability density function). The loops on the left (blue/red) have two crossings of the vertical axis (bistable probability density function), corresponding to XOR="FALSE". The unstable segments (dotted) can be obtained via the theory, however they are not realized in the experiments. See text for additional details.

isolate each tanh function developed by M1, M2, and M5. Note the currents of M1CMA & B ensure fully differential operation through the use of a common mode circuit (not shown). The common mode circuit maintains equal currents through M1CMA & B totaling the sum of tail currents through M5.

The general nature of each row lends extendability to the hardware implementation of large computational arrays as demonstrated by Figure 7 and Figure 8. The approach presented here can also be easily extended to arrays with backcoupling connections. As a direct result of this extendability, a simple cell library consisting of 30 diagonal and off-diagonal cells and their common mode cells will result in a powerful toolbox exploiting cell reuse in the fabrication of large weight arrays.

Figure 9 is a combination of a theoretical (from direct integration of the deterministic system) hysteresis loop (blue/green) and the loop (red/brown) obtained via the circuit. The red curve was generated directly from a Cadence simulation by incrementally adjusting the input current over the range $[-150, 150]$ nA while holding $\epsilon_4 = \epsilon_5 = 50$ nA; this yields the symmetric case where two "TRUE" inputs generate a "FALSE" output with the hysteresis curves almost symmetric about the vertical axis. The blue curves were generated by numerically integrating a deterministic version of equations (56), where an additional inhomogeneous forcing term was added to the equation for v_1. The input current is swept from -150 nA to 150 nA and back. The second hysteresis loop (on right–brown) is generated in a similar manner, with $\epsilon_4 = 50$ nA and $\epsilon_5 = -50$ nA; this is the case where one "TRUE" input and one "FALSE" input generate a "TRUE" output with its corresponding numerical integration case (green).

We note very good agreement between the theoretical and circuit results. These hysteresis loops should be viewed in the context of Figure 2. The (blue) switching curve, derived directly from the theory, is shown in the symmetric case (left set of loops in Figure 9) with its corresponding curve from the simulated circuit results (red). In this case, the blue and red switching curves cross the vertical axis in three places (three real roots). The hysteresis loops are nearly symmetrical about the vertical axis with the potential energy function (Equation 40) being nearly symmetric-bistable. The corresponding probability density function consists of two peaks of nearly equal area, centered at the points of intersection of the switching curves with the vertical axis. This is a case where the XOR function evaluates to "FALSE". The other set of hysteresis loops on the right portion of the figure (green/brown) where one input is "TRUE" and the other is "FALSE" display switching curves which will have only one intersection with the vertical axis (one real root and two complex conjugate roots); these hysteresis loops are shifted to the right. For this case, the probability density function is monomodal, the switching curve (not shown) intersects the vertical axis at a single point, and the XOR function evaluates to "TRUE". In both cases, the roots of the switching curves (that is, the locations of the extrema of the probability density function $P(y_1)$), as well as the y-axis crossings of the hysteresis loops, are nearly identical via the theory and its circuit realization. It is important to note that the dotted segment of the blue switching curve (shown in the symmetric case) is dynamically unstable. It can be derived from theory, however, it does not appear in the experimental data. For the red curves we obtain vertical jumps between the two steady states, as explained above, and the dotted red curve in the symmetric case has been synthetically added for comparison with the theoretical switching (blue) curve, to illustrate agreement between theory and experiment.

This agreement suggests that, given a set of theoretical parameters which point to a regime (e,g, in Figure 3 or Figure 4) wherein the XOR can be realized, a nonlinear circuit manifestation of the coupled system would generate the XOR for the parameters obtained from the theory (and carefully transformed to the corresponding circuit parameters).

V. DISCUSSION

A. UNIVERSAL FUNCTION APPROXIMATION

We have focused on the use of coupled nonlinear dynamic elements to produce a network that implements an XOR as a concrete example. Our networks have, in principle, the same representation power as standard ANNs with a sigmoid activation function. Focusing on networks with a layer of n input nodes, an internal layer of m nodes, and a single output node, an ANN implements a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, which can be written as

$$f(x) = \sum_{i=1}^{m} \alpha_i \sigma(\theta_i + y_i^t x). \tag{60}$$

Here, σ is the usual sigmoid activation function, and $\theta_i \in \mathbb{R}$, $y_i \in \mathbb{R}^n$ are the network biases and weights, respectively. It was shown in [27] that sums of the form (60) are dense in the space $C(I_n)$ of continuous functions on the unit hypercube $I_n = [0, 1]^n \subset \mathbb{R}^n$ with the uniform (supremum) norm. This means any continuous function $f : I_n \longrightarrow \mathbb{R}$ can be approximated arbitrarily closely by an ANN with a single hidden layer (albeit with m possibly very large). In this context, we reiterate that our theoretical approach of Section II can be applied to an arbitrary sized network and also incorporate back coupling when needed. This was, in fact, demonstrated by us in a simplified network ($N = 2$) in our previous work [12].

As described in Section II, a coupled network of dynamic elements with n input elements, m internal elements, and a single output element is essentially performing the computation

$$x \mapsto \alpha_0 + \sum_{i=1}^{m} \alpha_i \tanh\left(\theta_i + \sum_{j=1}^{n} \alpha_{ij} \tanh(\beta_{ij} x_j)\right). \tag{61}$$

While the summands of (61) are somewhat different from those in (60), it can be shown using the same technique as in [27] that sums of the form (61) are also dense in $C(I_n)$. Therefore, in principle, one can set up a dynamic network to approximate any desired computation. What is missing, however, is a systematic procedure for finding the right parameters, akin to training an ANN with backpropagation. Once such a procedure is established, one could foresee the ability to perform complex tasks, for example, pattern recognition on data sets such as images, where pixel RGB values are encoded as constant bias signals into the input layer.

B. POWER CONSUMPTION

At this stage, some remarks on power consumption in the nonlinear circuit in Section V are warranted. A classical figure of merit for low power analog circuits is the dynamic range (DR). Dynamic range is limited by the signal amplitude which in turn, is limited by the power supply (V_{DD}) and the system noise floor. For the system under consideration, the dynamic range (DR) can be written as [31]

$$\text{DR} = \frac{(V_{DD} - 20U_T)^2}{\kappa / I} \propto I V_{DD}^2 \tag{62}$$

where κ is a constant taken from the noise term in (51) and the $20U_T$ ensures all devices generating weight elements (see Figure 8) remain in their active analog regimes. The power consumption is proportional to DR [31], [32]

$$P = I\, V_{DD} \propto \frac{\text{DR}}{V_{DD}} \tag{63}$$

with equations (62) and (63) being strictly valid at low noise where the system gain is approximately linear.

Note that thermal noise, in the circuit presented here, is the driving force behind the sampling or changes of state. From the hysteresis loops in Figure 9, it is apparent that a change greater than 60 nA (from +30 nA to −30 nA or vice versa) is required to generate a change in state (i.e. a switch). This change of state represents a ±30 nA change in the output current I, which at 5σ, suggests that the noise standard deviation will be on the order of 12 nA. As a result, the DR can be defined as the state change divided by the noise standard deviation; in this specific example, one obtains a DR $= 20 \log(60/12) = 13.98$ dB. Therefore, the desirable DR should be approximately $10 - 16$ dB. Note that too large a noise signal can distort the system, resulting in unreliable results, whereas a noise signal that is too small will not produce a result at all. Hence, an optimal DR likely exists; however, determining its exact value for this system is beyond the scope of this study. This behavior is in contrast to classical analog systems, such as analog-to-digital converters, where the desired DR's are 60 dB to 150 dB or higher.

From a hardware perspective, there are many other hysteretic, nonlinear elements beyond CMOS circuits that are more susceptible to noise, and thus might be suitable for this method of computation. These include a range of nonlinear devices such as superconducting elements [28], memristors [29] and ferromagnetic nanowires [30]. Once the circuit is constructed using one of these devices as the nonlinearity, considerations such as the appropriate amount of noise necessary for the computation, determination of the coupling coefficients for a desired transform, etc., can be explored within this framework, using the theory as a guide.

All in all, the creation of an analog computing environment, due to a lower DR, is likely to produce computations that can accomplished with less power than traditional analog approaches, not to mention being able to function in noise environments unsuitable for classical digital computing. The application of this method is more suited to systems where the system switching is driven by the combination of the ‘‘hidden’’ digital signal and noise contained within the signal and power constrained hardware, e.g. quantum mechanical systems including spintronics, or systems based on SQUID devices.

C. NOISE

Since the noise floor has been an important component of the dynamics addressed in this paper, some general (noise-related) closing remarks are in order. The noise terms $N_i(t)$ in (1) are assumed to arise internally (e.g. thermal noise). However, in real-world applications there could also be external noise attached to an input signal (e.g., the dc inputs ε_i), and in some sensor applications of an array of the form (1) there would be other sources of noise. As an example of the latter, we can envision an array of coupled magnetic field sensors which would be susceptible to atmospheric magnetic noise arising from variations in the terrestrial magnetism, lightning strikes, seismic activity, etc. Environmental monitoring sensors are also susceptible to external noise, because signals to be monitored are, usually, aperiodic and could easily be overlapped by external broadband noise. The (external) noise effects usually have to be dealt with on a case-by-case basis, for example during design time, for characterization purposes. In the system studied in this paper, one could have an additional (additive) noise term in one or all of the equations; this leads to additional diffusion terms in the FPE (17). As long as the additional noise terms are additive, the ensuing FPE and its solution will have a similar structure to what we obtained in Section II. Their effects on the switching dynamics of the array and the realization of the XOR can be studied as a function of the different noise standard deviations; these effects are, of course, intimately connected to the nonlinearity and coupling strengths (i.e. the matrix J_{ik}). In this paper the noise terms in each equation are taken to represent a combination of internal noise, as well as fluctuations superimposed on the applied dc signals. For the sake of the phenomenology discussed in this paper, this is a sufficient treatment.

In a nonlinear device with readout dependent on switching between stable states, a reasonably large switch rate is desired to achieve reasonable sampling of the output in a short observation time. This is especially true when these devices are used in computation/logic circuits. In general, the switch rate is governed by the energy barrier height, applied external power, and noise floor intensity. In a coupled system, as described in this work, the effective energy barrier of the readout element is a function of the input signals, the coupling coefficients, and the noise floor standard deviation in each element, evident in our $N = 5$ system, from equations (40)-(42). Subsequently, the net power required to switch the readout element can be adjusted by adjusting these (deterministic) parameters. Adjusting them effectively adjusts the noise floor through their effects on the energy barrier; in turn, this changes the switching rate. This is analogous to the stochastic resonance scenario [11], wherein such an adjustment has been referred to as ‘‘tuning’’ to the noise. In our coupled system, the readout element can be made to operate efficiently at a lower dynamic range. Thus we can speculate that in systems/devices with larger noise floors, for example Resistive Random Access Memory (RRAM) [33] arrays, the background noise might significantly enhance the array performance, with careful parameter adjustments; however, too much noise could degrade performance.

VI. CONCLUSION

We have demonstrated the potential viability of a novel method of computation using a set of coupled noisy nonlinear elements, configured to yield XOR logic. This method represents a new way of generating neural-like dynamics, in the sense of both computational neuroscience and machine learning. Theoretical calculations and basic CMOS modeling of the system are in good agreement for the parameter regimes wherein the XOR can be realized. While scaling this technology up to practical use would require significant additional investment in fabrication technologies or simulation capacity, our results validate a powerful method of attacking a large set of problems. Optimization of the coupling parameters for any required scenario is virtually impossible through numerical simulations alone, due to the sheer number of parameter combinations available. However, with a theory that agrees with simulations and the circuit replication of our coupled system, such an optimization can be more efficiently achieved from a physically-informed starting point.

The overall good agreement of the theoretical results with the CADENCE simulations comes with an important caveat, however. CADENCE, while convenient, is slow, and is sometimes unable to faithfully reproduce all the desired circuit functionalities and characteristics, particularly when the circuit models contain complex dynamics. Other errors can arise from the circuit hardware not reproducing complex functions precisely. For instance, we know that the $\tanh(y)$ function is not properly reproduced in the circuit (the deviations occur mainly for large $\pm y$); this likely accounts for the small differences in the theoretical and experimental hysteresis loops in Figure 9.

The above remarks must be tempered by the observation that real-world circuit parameters are never exact; this is well-known in the circuit repertoire. This discrepancy can be addressed, theoretically, by allowing the system parameters (e.g. the dc inputs ε_i) (16) to have small deviations (with an appropriate distribution) about a mean value. Then, for example we might have $\varepsilon_i = \bar{\varepsilon_i} + \delta\varepsilon_i$ with the first term representing the deterministic value as used throughout this paper, and the second term representing small parameter deviations occurring in the circuit. At the end of the calculation, one would carry out an additional average over this distribution of the deviations. This process can get quite complicated/tedious if many adjustable parameters deviate in this manner from their expected values. A far more complicated situation can occur if the system parameters (e.g. the coupling coefficients) that are attached to functions of the dependent variable are actually affected by circuit noise leading to them being time-dependent and having their own correlation functions similar to (4). In this case one has multiplicative or "state-dependent" noise terms due to products of the fluctuating coupling coefficients and functions of the state variables y_i. The analytic treatment of Section II becomes far more complicated due to the FPE (17) having state-dependent diffusion terms (these are the second order terms on the rhs of (17) [23]). The above scenarios are, however, beyond the scope of this paper.

Finally, we be note that a neuromorphic system such as the one discussed in this work is inherently less sensitive to noise (or might actually take advantage of cooperative behavior mediated by the noise-floor), and can, in fact, be designed to operate in noisier environments. It is, therefore, reasonable to conjecture that our methodology may fit applications wherein the noise-floor might otherwise be detrimental to performance, better than traditional approaches. As just one example, we cite the case of wireless communication with drones where we could anticipate receiving noise contaminated data [34].

Further work will help to determine whether the highly distinct energy cost profiles associated with this method will enable truly low-power computation by taking advantage of its atypically reduced DR requirements, especially in environments with noisy signals or in technologies where the inherent noise is higher.

ACKNOWLEDGMENT

The authors would like to thank Prof. Subhasish Mitra with Stanford University for many helpful discussions.

REFERENCES

[1] A. Jafarian, P. Zeidman, V. Litvak, and K. Friston, "Structure learning in coupled dynamical systems and dynamic causal modelling," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 377, no. 2160, Dec. 2019, Art. no. 20190048, doi: 10.1098/rsta.2019.0048.

[2] I. Tsuda, "Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems," *Behav. Brain Sci.*, vol. 24, no. 5, pp. 793–810, Oct. 2001. [Online]. Available: https://kaylab.uchicago.edu/manuscripts/Tsuda_chaotic_itin_BBS.pdf

[3] M. Breakspear, "Dynamical models of large-scale brain activity," *Nature Neurosci.*, vol. 20, pp. 340–352, Mar. 2017. [Online]. Available: https://www.ngds-ku.org/Papers/J61/Breakspear.pdf

[4] C. Koch and J. Davis, Eds., *Large-Scale Neuronal Theories of the Brain*. Cambridge, MA, USA: MIT Press, 1994. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA290200.pdf

[5] M. Davies et al., "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, Jan. 2018.

[6] F. Broccard, S. Joshi, J. Wang, and G. Cauwenberghs, "Neuromorphic neural interfaces," in *Handbook of Neural Engineering*. Singapore: Springer, 2022, pp. 1–33.

[7] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1987.

[8] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[9] W. S. McCulloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943. [Online]. Available: https://jontalle.web.engr.illinois.edu/uploads/410-NS.F22/McCulloch-Pitts-1943-neural-networks-ocr.pdf

[10] W. Maas and H. Markram, "On the computational power of recurrent circuits of spiking neurons," *J. Comput. Syst. Sci.*, vol. 69, no. 4, pp. 593–616, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022000004000406

[11] A. Bulsara and L. Gammaitoni, "Tuning in to noise," *Phys. Today*, vol. 49, no. 39, pp. 39–47, 1996. [Online]. Available: https://scholar.archive.org/work/v7npagpf5jcatpm6fyhad3zu4y/access/wayback/http://www.nipslab.org/files/vol49no3p39-45.pdf

[12] M. E. Inchiosa and A. R. Bulsara, "Nonlinear dynamic elements with noisy sinusoidal forcing: Enhancing response via nonlinear coupling," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 52, no. 1, pp. 327–339, Jul. 1995, doi: 10.1103/PhysRevE.52.327.

[13] J. F. Lindner, B. K. Meadows, W. L. Ditto, M. E. Inchiosa, and A. R. Bulsara, "Array enhanced stochastic resonance and spatiotemporal synchronization," *Phys. Rev. Lett.*, vol. 75, no. 1, pp. 3–6, Jul. 1995. [Online]. Available: https://www.academia.edu/download/82000442/3bb611c2baa88d32c6e055743bd84ab7be2b.pdf

[14] K. Murali, S. Sinha, W. L. Ditto, and A. R. Bulsara, "Reliable logic circuit elements that exploit nonlinearity in the presence of a noise floor," *Phys. Rev. Lett.*, vol. 102, no. 10, Mar. 2009, Art. no. 104101. [Online]. Available: http://repository.ias.ac.in/60938/1/69-pub.pdf

[15] D. N. Guerra, A. R. Bulsara, W. L. Ditto, S. Sinha, K. Murali, and P. Mohanty, "A noise-assisted reprogrammable nanomechanical logic gate," *Nano Lett.*, vol. 10, no. 4, pp. 1168–1171, Apr. 2010. [Online]. Available: https://www.academia.edu/download/73892288/pdf.pdf

[16] K. Murali, S. Rajasekar, M. V. Aravind, V. Kohar, W. L. Ditto, and S. Sinha, "Construction of logic gates exploiting resonance phenomena in nonlinear systems," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 379, no. 2192, pp. 1–20, Mar. 2021, doi: 10.1098/rsta.2020.0238.

[17] A. Dari, B. Kia, X. Wang, A. R. Bulsara, and W. Ditto, "Noise-aided computation within a synthetic gene network through morphable and robust logic gates," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 83, no. 4, Apr. 2011, Art. no. 041909. [Online]. Available: https://www.academia.edu/download/90033689/acb30c1b430cb8bc4278767926d7d2787dd1.pdf

[18] A. Dari, B. Kia, A. R. Bulsara, and W. Ditto, "Creating morphable logic gates using logical stochastic resonance in an engineered gene network," *Europhys. Lett.*, vol. 93, no. 1, p. 18001, Jan. 2011, doi: 10.1209/0295-5075/93/18001.

[19] H. Haken, *Synergetics*. Berlin, Germany: Springer Verlag, 1983.

[20] R. Hasani, M. Lechner, A. Amini, L. Liebenwein, A. Ray, M. Tschaikowski, G. Teschl, and D. Rus, "Closed-form continuous-time neural networks," *Nature Mach. Intell.*, vol. 4, no. 11, pp. 992–1003, Nov. 2022.

[21] A. Karajgi. (2022). *How Neural Networks Solve the XOR Problem, Towards Data Science*. [Online]. Available: https://towardsdatascience.com/how-neural-networks-solve-the-xor-problem-59763136bdd7

[22] M. Hagan, H. Demuth, M. Beale, and O. de Jesus, *Neural Network Design*, 2nd ed. Stillwater, OK, USA: Oklahoma State Univ., 2014.

[23] C. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. Berlin, Germany: Springer Verlag, 1983.

[24] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of Brownian motion," *Phys. Rev.*, vol. 36, no. 5, p. 823, 1930. [Online]. Available: https://facultystaff.richmond.edu/~olipan/Uhlenbeck%20Orstein.pdf

[25] P. Kloeden and E. Platen, "Stochastic differential equations," in *Numerical Solution of Stochastic Differential Equations* (Applications of Mathematics), vol. 23. Berlin, Germany: Springer, doi: 10.1007/978-3-662-12616-5_4.

[26] R. Baker, *CMOS: Circuit Design, Layout, and Simulation* (IEEE Press Series on Microelectronic Systems), 4th ed. Hoboken, NJ, USA: Wiley, 2019.

[27] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control, Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989. [Online]. Available: https://hal.science/hal-03753170/document

[28] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect*. New York, NY, USA: Wiley, 1982.

[29] A. Thomas, "Memristor based neural networks," *J. Phys. D, Applied Phys.*, vol. 46, no. 9, 2013, Art. no. 093001. [Online]. Available: https://www.academia.edu/download/99354395/e671f7a69af730c3db0f7efa763860f7219a.pdf

[30] W. Geyger, *Nonlinear Magnetic Control Devices*. New York, NY, USA: McGraw-Hill, 1964.

[31] V. Baschirotto, G. Chironi, S. Cocciolo, M. D'Amico, P. Matteis, and P. Delizia, "Low power analog design in scaled technologies," in *Proc. Topical Workshop Electron. Part. Phys.*, 2009, pp. 103–110. [Online]. Available: https://cds.cern.ch/record/1234878/files/p103.pdf, doi: 10.5170/CERN-2009-006.103.

[32] A. Fahim, "Challenges in low-power analog circuit design for sub-28 nm CMOS technologies," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Aug. 2014, pp. 123–126, doi: 10.1145/2627369.2631639.

[33] F. Zahoor, T. Z. Azni Zulkifli, and F. A. Khanday, "Resistive random access memory (RRAM): An overview of materials, switching mechanism, performance, multilevel cell (MLC) storage, modeling, and applications," *Nanosc. Res. Lett.*, vol. 15, no. 1, p. 90, Dec. 2020. [Online]. Available: https://link.springer.com/article/10.1186/s11671-020-03299-9

[34] V. N. Swamy, P. Rigge, G. Ranade, A. Sahai, and B. Nikolic, "Network coding for high-reliability low-latency wireless control," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2016, pp. 138–144. [Online]. Available: http://people.eecs.berkeley.edu/~bora/Conferences/2016/WCNC16.pdf

PAUL R. DE LA HOUSSAYE (Senior Member, IEEE) received the B.S. degree in applied physics from the California Institute of Technology, Pasadena, CA, USA, in 1980, and the M.S. and Ph.D. degrees in applied physics from Stanford University, Stanford, CA, USA, in 1982 and 1988, respectively.

During the summers of 1980 and 1981, he was with Advanced Micro Devices, Sunnyvale, CA, USA. From 1988 to 1989, he held a postdoctoral research position with Centre Nationale de la Recherche Scientifique, Paris, France. Since 1990, he has been with the Naval Information Warfare Center Pacific. He is the author of over 50 publications in the open literature and two book chapters. He has 22 patents. His research interests include high-speed silicon-on-sapphire, SiGe and III–V devices, high-temperature superconductors, MEMS, photonics, statistics, quantum devices, and medical electronics. He has also been functioning as a government representative assisting DARPA and ONR, such as program development, technology transfer, proposal review, and contract oversight. Recently, he has been involved in research in the areas of artificial intelligence (AI) and neural networks.

Dr. de la Houssaye is or has been a member of the American Vacuum Society, the Electrochemical Society, the National Academy of Inventors, and Sigma XI.

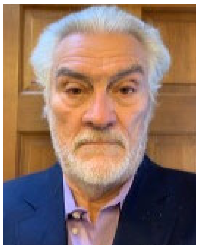
JUSTIN MAUGER received the B.S. degree in mathematics from the University of Washington, in 1995, and the Ph.D. degree in mathematics from Northwestern University, in 2001.

He has been with the Naval Information Warfare Center Pacific, San Diego, CA, USA, since 2013. Previously, he was a Senior Applications Engineer with Science Applications International Corporation. He is currently a Scientist with the Naval Information Warfare Center Pacific. His research interests include topological data analysis, machine learning, game theory, and neuromorphic computing.

ADI R. BULSARA received the B.Sc. degree in physics and mathematics from Bombay University, in 1970, the M.A. degree in physics from the State University of New York at Stony Brook, in 1973, and the Ph.D. degree in physics from The University of Texas at Austin, in 1978.

He holds the title of a Distinguished Scientist with the U.S. Naval Information Warfare Center Pacific, San Diego. He specializes in the applications of nonlinear dynamics in noisy systems, with an emphasis on nonlinear sensing, biomimetic sensors, stochastic resonance, noise-mediated cooperative behavior in sensor arrays, and nonlinear energy harvesting. In particular, he is recognized as one of the leading researchers on the stochastic resonance effect and its ramifications in coupled dynamical systems, including the physics of information processing in sensory neuron networks. In 1991, he (with A. Longtin and F. Moss) was the first to connect the inter-spike interval histogram (computed via biological experiments), with symmetries inherent in 2-state systems. Taken in conjunction with the definition of stochastic resonance as a bona fide resonance, this article postulated that information throughput in biological neurons could involve a stochastic resonance effect, mediated by the neuron noise floor, and when the input signal was weak. His review article "Tuning to the Noise" (with L. Gammaitoni) was the cover article in Physics Today (March 1996). He has been with the U.S. Naval Information Warfare Center Pacific, since 1983. He has authored over 200 articles in the physics and engineering literature and holds 25 patents.

Dr. Bulsara was a Welch Foundation Fellow with The University of Texas at Austin, in 1978. He is a fellow of the American Physical Society (APS) and the National Academy of Inventors (NAI). In 2017, he received the U.S. Navy's Highest Award (Fred Saalfeld Award) for a lifetime of outstanding achievements.



CHRISWELL HUTCHENS (Life Member, IEEE) received the B.S. and M.S. degrees from South Dakota State University and the Ph.D. degree in electrical engineering from the University of Missouri.

He has held employment with a successful startup (Daktronics, Brookings, SD, USA), in 1972, a Regional Cardiac Medical Center (73-78), Oklahoma State University, from 1982 to 2018, and most recently with NIWC Pacific. He has over 35 years' experience in analog SOI-CMOS design, research, and teaching. He holds nine patents. His research interests include extreme temperature and low power analog CMOS.

BENJAMIN MIGLIORI (Member, IEEE) received the B.S. and M.S. degrees in biophysics and the Ph.D. degree in physics/neuroscience from the University of California at San Diego, in 2005, 2008, and 2013, respectively.

After completing his dissertation in neuroscience, he joined the Space and Naval Warfare Systems Center, in 2013, where he developed and led research at the intersections of biological inspiration, neuroscience, and machine learning for the U.S. Navy. In 2018, he moved to the Los Alamos National Laboratory, Los Alamos, New Mexico, as a Staff Scientist. As a Senior Scientist with Los Alamos, he contributes to machine learning and neuromorphic computing for purposes of national defense. His work spans a wide range of technical levels, from basic science to engineering prototypes. Additionally, he contributes to the literature for novel adversarial machine learning methods.